

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:25:37 ; Search time 21 seconds

(without alignments)  
1204.398 Million cell updates/sec

Title: US-10-057-951-2

Perfect score: 1393

Sequence: 1 MLLAWVQAFVSNMLIAEAY.....PVDPQSGSTPLMGQAGTPGA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:\*

1: Pirl.\*

2: Pirl2.\*

3: Pirl3.\*

4: Pirl4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154.5	11.1	562	1 UKHUT	t-plasminogen acti
2	154	11.1	559	1 A35029	t-plasminogen acti
3	147.5	10.6	559	1 A29941	t-plasminogen acti
4	146.5	10.5	655	1 A46688	hepatocyte growth
5	145.5	10.4	291	2 I38098	t-plasminogen acti
6	142	10.2	810	1 PLHU	plasmin (EC 3.4.21
7	138	9.9	169	2 A40522	plasmin (EC 3.4.21
8	138	9.9	433	1 UKBAY	u-plasminogen acti
9	137	9.8	716	1 JC5061	macrophage-stimula
10	135.5	9.7	442	1 UKPG	u-plasminogen acti
11	134	9.6	431	2 JS0599	t-plasminogen acti
12	134	9.6	477	1 A34369	t-plasminogen acti
13	134	9.6	477	2 JS0598	t-plasminogen acti
14	134	9.6	716	1 A40332	macrophage-stimula
15	132	9.5	394	2 JS0600	t-plasminogen acti
16	132	9.5	431	1 UKHU	u-plasminogen acti
17	130	9.3	810	2 I46260	plasmin (EC 3.4.21
18	128.5	9.2	434	1 A35005	u-plasminogen acti
19	127	9.1	433	1 UN0560	u-plasminogen acti
20	126.5	9.1	728	1 A60185	hepatocyte growth
21	125.5	9.0	560	1 JC4795	plasma hyaluronan-
22	124.5	8.9	810	2 B30848	plasmin (EC 3.4.21
23	124	8.9	622	1 TBHU	thrombin (EC 3.4.2
24	124	8.9	728	1 JH0579	hepatocyte growth
25	124	8.9	812	1 FLBO	plasmin (EC 3.4.21
26	123	8.8	593	2 S45281	coagulation factor
27	122.5	8.8	458	1 S00657	apoptein(a) (EC
28	122	8.8	603	2 S28941	coagulation factor
29	121.5	8.7	728	1 A35644	hepatocyte growth

30	120	8.6	433	1 UKMS	u-plasminogen acti
31	120	8.6	790	1 PLPG	plasmin (EC 3.4.21
32	119.5	8.6	477	2 JS0597	t-plasminogen acti
33	119	8.5	710	1 I51283	hepatocyte growth
34	118.5	8.5	615	1 KFHU12	coagulation factor
35	118	8.5	432	1 S18932	u-plasminogen acti
36	117	8.4	625	1 TBBO	thrombin (EC 3.4.2
37	116	8.3	1420	2 A32869	apolipoprotein(a)
38	115.5	8.3	711	1 A47136	macrophage-stimula
39	114.5	8.2	812	1 PLMS	plasmin (EC 3.4.21
40	114	8.2	618	2 A35827	thrombin (EC 3.4.2
41	113	8.1	558	2 JC5878	plasma hyaluronan-
42	112	8.0	455	2 A61545	plasmin (EC 3.4.21
43	110.5	7.9	685	1 A48289	neurotrophic recep
44	108.5	7.8	123	2 C61545	plasmin (EC 3.4.21
45	107	7.7	617	2 S10511	thrombin (EC 3.4.2

#### ALIGNMENTS

##### RESULT 1

###### UKHUT

t-plasminogen activator (EC 3.4.21.68) precursor [validated] - human  
N;Alternate names: t-PA; tissue plasminogen activator  
C;Species: Homo sapiens (man)  
C;Date: 14-Nov-1983 #sequence\_revision 14-Nov-1983 #text\_change 08-Dec-2000  
C;Accession: A94004; A23529; J0562; A93293; S02125; A91343; A93951; A91322; A54645; I.  
R;NY, T.; Elgh, F.; Lund, B.  
Proc. Natl. Acad. Sci. U.S.A. 81, 5355-5359, 1984  
A;Title: The structure of the human tissue-type plasminogen activator gene: correlation:  
A;Reference number: A94004; MUID:84298137; PMID:6089198  
A;Accession: A94004  
A;Molecule type: DNA  
A;Residues: 1-562 <NYT>  
A;Cross-references: GB:L00141  
A;Note: the codon given for residue 93 (ACC) is inconsistent with the authors' transla  
B;Fierzner Degen, S.J.; Rajput, B.; Reich, B.  
J. Biol. Chem. 261, 6972-6985, 1986  
A;Title: The human tissue plasminogen activator gene.  
A;Reference number: A23529; MUID:86196143; PMID:3009482  
A;Accession: A23529  
A;Molecule type: DNA  
A;Residues: 1-562 <DEG>  
A;Cross-references: GB:X03021; NID:g339817; PIDN:AAA98809.1; PID:g339818  
R;Itagaki, Y.; Yasuda, H.; Morinaga, T.; Mitsuda, S.; Higashio, K.  
Agric. Biol. Chem. 55, 1225-1232, 1991  
A;Title: Purification and characterization of tissue plasminogen activator secreted by  
A;Reference number: J0562; MUID:91291340; PMID:1368681  
A;Accession: J0562  
A;Molecule type: mRNA  
A;Residues: 31-562 <ITA>  
A;Cross-references: DDBJ:D01096; NID:g220128; PIDN:BAA00881.1; PID:g441174  
A;Experimental source: embryonic lung fibroblast IMR-90 cells  
A;Note: part of this sequence, including the amino end of the mature protein, was conf  
R;Pennica, D.; Holmes, W.E.; Kohr, W.J.; Harkins, R.N.; Vehar, G.A.; Ward, C.A.; Benne-  
Nature 301, 214-221, 1983  
A;Title: Cloning and expression of human tissue-type plasminogen activator cDNA in Bsc:  
A;Reference number: A93293; MUID:83115262; PMID:6337343  
A;Accession: A93293  
A;Molecule type: mRNA  
A;Residues: 1-562 <PEN>  
A;Cross-references: GB:L00141  
A;Experimental source: melanoma cells  
R;Sasaki, H.; Saito, Y.; Hayashi, M.; Otsuka, K.; Niwa, M.  
Nucleic Acids Res. 16, 5695, 1988  
A;Title: Nucleotide sequence of the tissue-type plasminogen activator cDNA from human  
A;Reference number: S02125; MUID:88262579; PMID:3133640  
A;Accession: S02125  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-562 <SAS>  
A;Cross-references: EMBL:X07393; NID:g37243; PIDN:CAA30302.1; PID:g37244

A:Experimental source: fetal lung cells  
 R:Kagatani, H.; Tagawa, M.; Hatanaka, K.; Ikari, T.; Saito, A.; Bando, H.; Okada, K.; Ma  
 FBS Lett. 189, 145-149, 1985  
 A:Title: Expression in *Escherichia coli* of finger-domain lacking tissue-type plasminogen  
 A:Reference number: A91343; MUID:85285620; PMID:3896853  
 A:Accession: A91343  
 A:Molecule type: mRNA  
 A:Residues: 1-38, 'G', 86-433, 'E', 435-562 <KAG>  
 A:Experimental source: Detroit 562 cells; ATCC 138  
 R:Edlund, T.; Ny, T.; Ranby, M.; Heden, L.O.; Palm, G.; Holmgren, E.; Josephson, S.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 349-352, 1983  
 A:Title: Isolation of cDNA sequences coding for a part of human tissue plasminogen activator  
 A:Reference number: A93951; MUID:83169656; PMID:6572897  
 A:Accession: A93951  
 A:Molecule type: mRNA  
 A:Residues: 251-358 <EDL>  
 A:Experimental source: melanoma cells  
 R:Pohl, G.; Kallstrom, M.; Bergsdorf, N.; Wallen, P.; Jornvall, H.  
 Biochemistry 23, 3701-3707, 1984  
 A:Title: Tissue plasminogen activator: peptide analyses confirm an indirectly derived amino  
 A:Reference number: A90488; MUID:85000468; PMID:6433976  
 A:Contents: annotation: melanoma cells, partial sequence of residues 36-562, active and  
 R:Pohl, G.; Kaplan, L.; Einarsson, M.; Wallen, P.; Jornvall, H.  
 FEBS Lett. 168, 29-32, 1984  
 A:Title: Differences between uterine and melanoma forms of tissue plasminogen activator.  
 A:Reference number: A91322; MUID:84158956; PMID:6538514  
 A:Accession: A91322  
 A:Molecule type: protein  
 A:Residues: 33-45; 311-320 <POH>  
 A:Experimental source: uterus  
 A:Note: In the uterus, cleavage of the activation peptide may also occur after 38-Gln  
 R:van Zonneveld, A.J.; Veerman, H.; Pannekoek, H.  
 J. Biol. Chem. 261, 14214-14218, 1986  
 A:Reference number: A37567; MUID:87033611; PMID:3021732  
 A:Contents: annotation; fibrin binding site  
 R:Verheijen, J.H.; Caspers, M.P.M.; Chang, G.T.G.; de Munk, G.A.W.; Pouwels, P.H.; Enger  
 EMBO J. 5, 3525-3530, 1986  
 A:Title: Involvement of finger domain and kringle 2 domain of tissue-type plasminogen ac  
 A:Reference number: A37568; MUID:87161761; PMID:3030730  
 A:Contents: annotation; fibrin binding site  
 R:Dodd, I.; Nunn, B.; Robinson, J.H.  
 Thromb. Haemost. 59, 523-528, 1988  
 A:Title: Isolation, identification and pharmacokinetic properties of human tissue-type B  
 A:Reference number: A60902; MUID:89044681; PMID:3142086  
 A:Contents: annotation; novel forms of expressed recombinant t-PA  
 R:Haris, T.J.R.; Patel, T.; Marston, F.A.O.; Little, S.; Emtage, J.S.; Opdenakker, G.;  
 Mol Biol. Med. 3, 279-292, 1986  
 A:Title: Cloning of cDNA coding for human tissue-type plasminogen activator and its expr  
 A:Reference number: A54645; MUID:86284200; PMID:3090401  
 A:Accession: A54645  
 A:Molecule type: mRNA  
 A:Residues: 1-562 <HAR>  
 A:Cross-references: GB:M15518; NID:q190031; PIDN:AAA60111.1; PID:q190032  
 A:Note: Parts of this sequence were confirmed by peptide sequencing  
 R:Reddy, V.B.; Garramone, A.J.; Sasak, H.; Wei, C.  
 DNA 6, 461-472, 1987  
 A:Title: Expression of human uterine tissue-type plasminogen activator in mouse cells us  
 A:Reference number: I60110; MUID:88054470; PMID:2824147  
 A:Accession: I60110  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-562 <RES>  
 A:Cross-references: GB:M18182; NID:G340176; PIDN:AAA36800.1; PID:G340177  
 R:Fisher, R.; Waller, E.K.; Grossi, G.; Thompson, D.; Tizard, R.; Schleuning, W.D.  
 J. Biol. Chem. 260, 11223-11230, 1985  
 A:Title: Isolation and characterization of the human tissue-type plasminogen activator s  
 A:Reference number: I55232; MUID:85289338; PMID:3161893  
 A:Accession: I55232  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-36 <RE2>  
 A:Cross-references: GB:M11890; NID:G339837; PIDN:AAA61213.1; PID:G339839

C:Comment: Cleavage by plasmin or trypsin produces two chains held together by a single  
 C:Comment: t-PA converts plasminogen to plasmin by hydrolyzing a single Arg-Val bond.  
 C:Comment: t-PA binds chain A of fibrin by kringle 2 and the fibronectin type I repeat  
 C:Genetics:  
 A:Gene: GDB:PLAT  
 A:Cross-references: GDB:119496; OMIM:173370  
 A:Map position: 8p12-8p12  
 A:Introns: 24/3; 39/1; 85/1; 122/1; 180/2; 211/1; 268/2; 297/1; 362/2; 408/1; 455/3; 5  
 C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat;  
 C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; plasma; serine proteinase  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-32/Domain: propeptide #status predicted <PRO>  
 F:33-562/Product: t-plasminogen activator #status experimental <MAT>  
 F:33-310/Product: t-plasminogen activator chain A #status experimental <ACH>  
 F:41-78/Domain: fibronectin type I repeat homology <1F1>  
 F:86-119/Domain: EGF homology <EGF>  
 F:127-208/Domain: kringle homology <KR1>  
 F:215-296/Domain: kringle homology <KR2>  
 F:311-562/Product: t-plasminogen activator chain B #status experimental <BCH>  
 F:311-556/Domain: trypsin homology <TRY>  
 F:41-71,69-78,86-97,91-108,110-119,127-208,148-190,179-203,215-296,236-278,267-291,299  
 F:152,483/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:219/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental  
 F:310-311/Cleavage site: Arg-Ile (plasmin, trypsin) #status experimental  
 F:357,406/Active site: His, Asp #status predicted  
 F:513/Active site: Ser #status experimental  
 Query Match 11.1%; Score 154.5; DB 1; Length 562;  
 Best Local Similarity 39.6%; Pred. No. 2.9e-05;  
 Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;  
 QY 25 CFWNGHLYREDQTSAPGLRCLNMLDQAGSLAPVS-----GAGNHSYCRNFE 75  
 Db 127 CYEDQGISYRGTSWTSAGSGAECTNW--NSSALAQPKYSGRPDALRLGLGNHNYCRNDR 184  
 QY 76 DPRGPWCYVSGEAGVPEKRPCEDLRCPTTS 106  
 Db 185 DSK-PWCYVF-KAGKYSEFCSTPACSEGN 213  
 RESULT 2  
 A35029  
 C:plasminogen activator (EC 3.4.21.68) precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A35029; A31597  
 R:Peng, P.; Ohlsson, M.; Ny, T.  
 J. Biol. Chem. 265, 2022-2027, 1990  
 A:Title: The structure of the TARA-less rat tissue-type plasminogen activator gene. S  
 A:Reference number: A35029; MUID:90130448; PMID:2105315  
 A:Accession: A35029  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-559 <PEN>  
 A:Cross-references: GB:M31197; NID:g207429; PIDN:AAA42261.1; PID:g207431; GB:J05226  
 R:Ny, T.; Leonardsson, G.; Hsueh, A.J.W.  
 DNA 7, 671-677, 1988  
 A:Title: Cloning and characterization of a cDNA for rat tissue-type plasminogen activator  
 A:Reference number: A31597; MUID:89170114; PMID:3148445  
 A:Accession: A31597  
 A:Molecule type: mRNA  
 A:Residues: 1-379, 'K', 381-559 <NYT>  
 A:Cross-references: GB:M23697; NID:g530159; PIDN:AAA41812.1; PID:g530160  
 C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat;  
 C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
 F:1-17/Domain: signal sequence #status predicted <SIG>  
 F:18-29/Domain: propeptide #status predicted <PRO>  
 F:30-308/Product: t-plasminogen activator #status predicted <MAT>  
 F:38-75/Domain: fibronectin type I repeat homology <1F1>  
 F:83-116/Domain: EGF homology <EGF>  
 F:124-205/Domain: kringle homology <KR1>  
 F:213-294/Domain: kringle homology <KR2>

F:309-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
F:309-553/Domain: trypsin homology <TRY>  
F:38-68,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4  
F:149,481/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted  
F:355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 11.1%; Score 154; DB 1; Length 559;  
Best Local Similarity 32.4%; Pred. No. 3.2e-05;  
Matches 45; Conservative 14; Mismatches 51; Indels 26; Gaps 6;

Qy 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLGASAPVS-----GAGNHSYCRNPDE 75  
Db 124 CFEGQIITYRGTSWTASGAECINW--NSSLSQKPYARRPNAIKLGLGNHNYCRNPDR 181

Qy 76 DPROPCWYVSGEAGVPEKRPCEPDLRCPD-----TTSQALPAPTTEIQBASEGPG 124  
Db 182 DVK-PWCYVF-KAGKYTTEFCSTPACPKGPTDCVVGKGYRGTGTHSFTT--SKASCLPW 237

Qy 125 ADEVQVFAPNALPARSEA 143  
Db 238 NSMILIGTKYTWANRANSOA 256

RESULT 3  
A29941  
t-plasminogen activator (EC 3.4.21.68) precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
A:Accession: A29941; S48205; S48206  
R:Ricklefs, R.J.; Darrow, A.L.; Strickland, S.  
J. Biol. Chem. 263, 1563-1569, 1988  
A:Title: Molecular cloning of complementary DNA to mouse tissue plasminogen activator mRNA  
A:Reference number: A29941; MUID:88087303; PMID:2826484  
A:Accession: A29941  
A:Molecule type: mRNA  
A:Residues: 1-559 <RIC>  
A:Cross-references: GB:J03520; NID:G202109; PIDN:AAA40470.1; PID:G202110  
R:Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.  
Eur. J. Biochem. 224, 863-871, 1994  
A:Title: Characterization of the murine plasma fibrinolytic system.  
A:Reference number: S48202; MUID:95010076; PMID:7523120  
A:Accession: S48205  
A:Molecule type: protein  
A:Residues: 33-37,'X',39-40 <LIJ>  
A:Accession: S48207  
A:Molecule type: protein  
A:Residues: 309-316 <LI2>  
A:Accession: S48206  
A:Molecule type: protein  
A:Residues: 33-37,'X',39-40 <LIW>  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat homology  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; signal sequence  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-29/Domain: propetide #status predicted <PRO>  
F:30-559/Product: t-plasminogen activator #status predicted <MAT>  
F:30-308/Product: t-plasminogen activator chain A #status predicted <ACH>  
F:83-116/Domain: fibronectin type I repeat homology <1FI>  
F:83-124/Domain: EGF homology <EGF>  
F:124-205/Domain: kringle homology <KR1>  
F:213-294/Domain: kringle homology <KR2>  
F:309-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
F:309-553/Domain: trypsin homology <TRY>  
F:38-68,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4  
F:149,481/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted  
F:355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 10.6%; Score 147.5; DB 1; Length 559;  
Best Local Similarity 37.0%; Pred. No. 0.00011;  
Matches 34; Conservative 11; Mismatches 34; Indels 13; Gaps 4;

Qy 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLGASAPVS-----GAGNHSYCRNPDE 75

Db 124 CFEGQIITYRGTSWTASGAECINW--NSSLSKPYARRPNAIKLGLGNHNYCRNPDR 181  
Qy 76 DPROPCWYVSGEAGVPEKRPCEPDLRCPD-----TTSQ 107  
Db 182 DLK-PWCYVF-KAGKYTTEFCSTPACPKGKSE 211

RESULT 4  
A46688  
hepatocyte growth factor activator (EC 3.4.21.-) precursor [validated] - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Sep-1993 #sequence\_revision 25-Aug-1995 #text\_change 08-Dec-2000  
A:Accession: A46688  
R:Myazawa, K.; Shimomura, T.; Kitamura, A.; Kondo, J.; Morimoto, Y.; Kitamura, N.  
J. Biol. Chem. 268, 10024-10028, 1993  
A:Title: Molecular cloning and sequence analysis of the cDNA for a human serine protease d coagulation factor XII.  
A:Reference number: A46688; MUID:93252878; PMID:7683665  
A:Accession: A46688  
A:Molecule type: mRNA  
A:Residues: 1-655 <MIY>  
A:Cross-references: DDBJ:D14012; NID:G219680; PIDN:BAAG03113.1; PID:G219681  
A:Experimental source: liver (mRNA); serum (protein)  
A:Note: sequence extracted from NCBI backbone (NCBIN:131227, NCBIPI:131228)  
A:Note: parts of the sequence, including the amino ends of the heavy and light chains.  
C:Genetics:  
A:Gene: GDB:HGFAC; HGFA; HGFAP  
A:Cross-references: GDB:9954514  
A:Map position: 4p16-4p16  
C:Function:  
A:Description: activates hepatocyte growth factor by specific proteolytic cleavage  
A:Pathway: tissue repair and regeneration  
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology  
C:Keywords: glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:108-148/Domain: fibronectin type II repeat homology <1F2>  
F:164-197/Domain: EGF homology <EG1>  
F:202-237/Domain: fibronectin type I repeat homology <1FI>  
F:245-278/Domain: EGF homology <EG2>  
F:286-367/Domain: kringle homology <KRG>  
F:373-407/Product: hepatocyte growth factor activator light chain #status experimental  
F:408-655/Product: hepatocyte growth factor activator heavy chain #status experimental  
F:408-641/Domain: trypsin homology <TRY>  
F:40,48,290,468,492,546/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:164-175,169-186,188-197,202-230,228-237,245-256,250-267,269-278,286-367,307-349,338-  
F:447,497,598/Active site: His, Asp, Ser #status predicted

Query Match 10.5%; Score 146.5; DB 1; Length 655;  
Best Local Similarity 36.9%; Pred. No. 0.00017;  
Matches 41; Conservative 10; Mismatches 41; Indels 19; Gaps 6;

Qy 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSLGASAPVSAGNHSYCRNPDE 76  
Db 286 CFLNGTGYRGVASTASGSLCLAWNSDLYLQELHVDV-VGAALLGLGPHAYCRNPND 344

Qy 77 PRGPCWYVSGEAGVPEKRPCEPDLRCPD-----EKRPCEPDLRCPDPAFTTEIQE-ASEG 122  
Db 345 ER-PWCYVVKDSALSWBYCRLEACESL---TRVQLSPDLLATPEPASPG 390

RESULT 5  
I38098  
t-plasminogen activator precursor, inactive endothelial splice form - human  
N:Alternate names: tissue plasminogen activator  
C:Species: Homo sapiens (man)  
C:Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 22-Jun-1999  
A:Accession: I38098; S01678  
R:Siebert, P.D.; Fong, K.  
Nucleic Acids Res. 18, 1086, 1990  
A:Title: Variant tissue type plasminogen activator (PLAT) cDNA obtained from human end  
A:Reference number: I38098; MUID:90192128; PMID:1969145  
A:Accession: I38098

A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-291 <SIG>  
A:Cross-references: EMBL:X13097; NID:g35282; PIDN:CAA31489.1; PFD:g35283  
C:Comment: For the main splice form, see PIR:UKHUT. This form probably does not have propeptides.  
C:Genetics:  
A:Gene: GDB:PLAT  
A:Cross-references: GDB:l19496; OMIM:173370  
A:Map position: 9p12-8p12  
A:Introns: 24/3; 39/1; 85/1; 122/1; 180/2; 211/1; 268/2  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat homologous domain  
C:Keywords: alternative splicing; fibrinolysis; glycoprotein; kringle  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-32/Domain: propetide #status predicted <PRO>  
F:33-291/Product: t-plasminogen activator, inactive endothelial splice form #status predicted <K2>  
F:41-78/Domain: fibronectin type I repeat homology <R1>  
F:86-119/Domain: EGF homology <EGF>  
F:127-208/Domain: kringle homology <KR1>  
F:215-291/Domain: kringle homology #status atypical <KR2>  
F:41-71,69,78,86-97,91-108,110-119,127-208,148-190,179-203/Disulfide bonds: #status predicted <FA>

Query Match            10.4%; Score 145.5; DB 2; Length 291;  
Best Local Similarity   38.5%; Pred. No. 7.9e-05;  
Matches         35; Conservative      6; Mismatches    37; Indels    13; Gaps          4;

DQ         25 CFWDNGHYLYREDQTSPAPGRCLNLWDAQSGLASAPVS-----GAGNHGYCRNPDE 75  
                 : : : : : : : : : : : : : : : : |  
Db         127 CYEDQGISTYRGVSTAESGAECTNW--NSSALAQNAVSGRRPDIAIRLGHNHCYNRPDR 184

DQ         76 DPRGPWCYVGSEAGVPKKRCEDLRCPETTS 106  
                 - : : : : : : : : : : : : : : : : |  
Db         185 DSK-PWCYVF-KAGKYSEFCSTPACSEGNS 213

RESULT 6

PLHU  
plasmin (EC 3.4.21.7) precursor [validated] - human  
N:Alternate names: plasminogen precursor [misnomer]  
N:Contains: angiotensin; microplasma; plasminogen  
C:Species: Homo sapiens (man)  
C>Date: 24-Apr-1984 #sequence revision 02-Dec-1994 #text change 15-Sep-2000  
C:Accession: A35229; F52242; A26646; J62738; I64609; S03735; A00929; A04627; A04625; A04626  
J.Petersen, T.E.; Martzen, M.R.; Ichinose, A.; Davie, E.W.  
J. Biol. Chem. 265, 6104-6111, 1990  
A>Title: Characterization of the gene for human plasminogen, a key proteolytic enzyme in the blood coagulation cascade  
A:Reference number: A35229; PMID:90202879; PMID:2318848  
A:Accession: A35229  
A:Molecule type: DNA  
A:Residues: 1-810 <PEPT>  
A:Cross-references: GB:J05286; GB:M34276; NID:G190064; PIDN:AAA60113.1; PFD:G387026  
A:Experimental source: leukocyte; lung fibroblast  
R:Malgaroli, N.; Bruno, L.; Pontoglio, M.; Candiani, G.; Meroni, G.; Ottolenghi, S.; Tarantini, C.; Vignani, R.; Bazzucchi, L.; Santoro, S.G.  
Biochem. Biophys. Res. Commun. 173, 1013-1018, 1990  
A>Title: Definition of the transcription initiation site of human plasminogen gene in liver tissue  
A:Reference number: F52242; PMID:91097523; PMID:2268308  
A:Accession: F52242  
A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-16 <MALL>  
A:Cross-references: GB:M62890; NID:G190092; PIDN:AAA36454.1; PFD:G553613  
R:Forsgren, M.; Radén, B.; Israelsson, M.; Larsson, K.; Heden, L.O.  
FEBS Lett. 213, 254-260, 1987  
A>Title: Molecular cloning and characterization of a full-length cDNA clone for human plaminogen activator-related proteinase  
A:Reference number: A26646; PMID:87162490; PMID:3030813  
A:Accession: A26646  
A:Molecule type: mRNA  
A:Residues: 1-471,'D',473-810 <FOR>  
A:Cross-references: GB:X05199; NID:G35530; PIDN:CAA28831.1; PFD:G35531  
A:Experimental source: liver  
R:Mainowski, D.P.; Sadler, J.E.; Davie, E.W.  
Biochemistry 23, 4243-4250, 1984  
A>Title: Characterization of a complementary deoxyribonucleic acid coding for human and rat plasminogen activator-related proteinase  
A:Reference number: I45961; PMID:85023311; PMID:6148961



A:Reference number: A58812; PMID:9548733; PMID:9548733  
A:Contents: annotation  
R:Tulinsky, A.; Mulichak, A.M.  
submitted to the Brookhaven Protein Data Bank, July 1991  
A:Reference number: A51341; PDB:1PK4  
A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 376-454  
R:Tulinsky, A.; Wu, T.P.  
submitted to the Brookhaven Protein Data Bank, July 1991  
A:Reference number: A51488; PDB:2PK4  
A:Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 375-454  
R:Wu, T.P.; Tulinsky, A.  
submitted to the Brookhaven Protein Data Bank, August 1993  
A:Reference number: A51911; PDB:1PKR  
A:Contents: annotation; X-ray crystallography, 2.48 angstroms, residues 102-181  
R:Padmanabhan, K.; Tulinsky, A.  
submitted to the Brookhaven Protein Data Bank, April 1994  
A:Reference number: A52408; PDB:1PMK  
A:Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 377-454  
R:Tulinsky, A.; Mathews, I.I.  
submitted to the Brookhaven Protein Data Bank, December 1995  
A:Reference number: A65244; PDB:1CEA  
A:Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181  
R:Tulinsky, A.; Mathews, I.I.  
submitted to the Brookhaven Protein Data Bank, December 1995  
A:Reference number: A65245; PDB:1CEB  
A:Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181  
R:Mulichak, A.M.; Tulinsky, A.; Ravichandran, K.G.  
Biochemistry 30, 10576-10588, 1991  
A:Title: Crystal and molecular structure of human plasminogen kringle 4 refined at 1.9 A  
A:Reference number: A58819; PMID:92031502; PMID:1657148  
A:Contents: annotation  
R:Wu, T.P.; Padmanabhan, K.; Tulinsky, A.; Mulichak, A.M.  
Biochemistry 30, 10589-10594, 1991  
A:Title: The refined structure of the epsilon-aminocaproic acid complex of human plasminogen  
A:Reference number: A58818; PMID:92031503; PMID:1657149  
A:Contents: annotation  
R:de Vos, A.M.; Ultsch, M.H.; Kelley, R.F.; Padmanabhan, K.; Tulinsky, A.; Westbrook, M.  
Biochemistry 31, 270-279, 1992  
A:Title: Crystal structure of the kringle 2 domain of tissue plasminogen activator at 2.2 A  
A:Reference number: A39483; PMID:92118803; PMID:1310033  
A:Contents: annotation; X-ray crystallography, 2.4 angstroms  
R:Stec, B.; Teeter, M.M.; Whitlow, M.; Yamano, A.  
submitted to the Brookhaven Protein Data Bank, June 1995  
A:Reference number: A65980; PDB:1KRN  
A:Contents: annotation; X-ray crystallography, 1.67 angstroms, residues 376-454  
R:Rejante, M.; Llinas, M.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A:Reference number: A65803; PDB:1HPJ  
A:Contents: annotation; conformation by (1)H-NMR, residues 103-181  
R:Rejante, M.; Llinas, M.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A:Reference number: A65804; PDB:1HPK  
A:Contents: annotation; conformation by (1)H-NMR, residues 103-181  
R:Rejante, M.R.; Llinas, M.  
Eur. J. Biochem. 221, 927-937, 1994  
A:Title: (1)H-NMR assignments and secondary structure of human plasminogen kringle 1.  
A:Reference number: A43645; PMID:94237157; PMID:8181475  
A:Contents: annotation; conformation by (1)H-NMR, residues 96-184  
R:Rejante, M.R.; Llinas, M.  
Eur. J. Biochem. 221, 939-949, 1994  
A:Title: Solution structure of the epsilon-aminohexanoic acid complex of human plasminogen  
A:Reference number: A58817; PMID:94237158; PMID:8181476  
A:Contents: annotation; conformation by (1)H-NMR  
C:Comment: Plasminogen is synthesized by the kidney and is present in plasma and many other tissues.  
C:Comment: Plasminogen is converted to plasmin by plasminogen activators (see PIR:UKHU a d PIR:FGHUB).  
C:Comment: Plasmin is inactivated by alpha-2-antiplasmin (see PIR:ITHUA2) immediately after release from the fibrinolytic system.  
C:Comment: Plasmin is formed by autolytic cleavage of plasminogen under artificial conditions.  
C:Comment: Streptolysin 1 (see PIR:KCHU1) acts on plasminogen to produce angiotensin. To inhibit this process, angiotensin-converting enzyme inhibitors are used.  
C:Gene: GDB:PLG

A:Cross-references: GDB:119498; OMIM:173350  
A:Map position: 6226-6q27  
A:Introns: 17/1; 62/2; 98/2; 136/2; 183/1; 223/2; 366/1; 419/2; 480/1; 525/2  
C:Function:  
A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a number of the walls of the graafian follicle; also activates the urokinase-type plasminogen activator; fibrinolysis  
A:Pathway: fibrinolysis  
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homologous  
C:Keywords: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hyaluronidase  
F:1-96/Domain: plasminogen-related protein precursor homology <PLPH>  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-810/Product: plasminogen #status experimental <PRO>  
F:20-96/Domain: activation peptide #status experimental <APT>  
F:79-466/Product: angiotensin #status experimental <AST>  
F:97-580,581-810/Product: plasmin #status experimental <MAT>  
F:97-580/Domain: plasmin chain A #status experimental <CHA>  
F:103-181/Domain: kringle homology <KR1>  
F:185-262/Domain: kringle homology <KR2>  
F:275-352/Domain: kringle homology <KR3>  
F:377-454/Domain: kringle homology <KR4>  
F:481-560/Domain: kringle homology <KR5>  
F:550-580,581-810/Product: microplasmin #status experimental <MMT>

Query Match 10.2%; Score 142; DB 1; Length 810;  
Best Local Similarity 37.2%; Pred. No. 0.00051;  
Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4;

QY 25 CFWDNGHLYREDQTSAPGLRCLNWLDA-----QSGLASAPVSGAGNHSCRNPDPRG 79  
DB 103 CXTGNKKNRYGTMTSKTKNGITCQKWSSTSPRRPSPATHPSEGL-EENYCRNPNDPOG 161  
||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :  
DQ 80 PWCYVSGEAGVPEKK--PCEDLRCPPE 103  
162 PWCYTID---PEKRYDYCDILECEE 183

RESULT 7  
A40522  
Plasmin (EC 3.4.21.7) precursor - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 28-Feb-1992 #sequence\_revision 17-Apr-1993 #text\_change 16-Jul-1999  
C:Accession: A40522  
J:Kanalas, J.J.; Makker, S.P.  
A:Title: Identification of the rat Heymann nephritis autoantigen (GP330) as a receptor for plasminogen activator  
A:Reference number: A40522; MUID:91250378; PMID:1645711  
A:Accession: A40522  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-169 <KAN>  
A:Cross-references: GB:M62832; NID:G206215; PIDN:AAA41884.1; PID:G554488  
A>Note: the authors translated the codon TCT for residue 76 as Ala  
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homologous  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:34-112/Domain: kringle homology <KRG>  
F:34-112,55-95,83-107/Disulfide bonds: #status predicted

Query Match 9.9%; Score 138; DB 2; Length 169;  
Best Local Similarity 31.8%; Pred. No. 0.00018;  
Matches 35; Conservative 13; Mismatches 48; Indels 14; Gaps 5;

QY 25 CFWDNGHLYREDQTSAPGLRCLNWLDA-----QSGLASAPVSGAGNHSCRNPDPRG 79  
DB 34 CYQGNGSKSYRTSSTNTGKCQSVMTSHSKTPANFPDSGL-EMNYCRNPNDQRG 92  
||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :  
QY 80 PWCYVSGEAGVPEKK--PCEDLRCPETTSQALPAFTTIQEASEGFDAE 127  
||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :  
DB 93 PWCFTTD---PSVRWEYCNLKRCSETGGV--AESAIVPQVPSAPGTSE 136

RESULT 8  
UKBAY  
plasminogen activator (EC 3.4.21.73) precursor - yellow baboon





```
Db      80  LLPWQTQ-----HSLHTQLYHSSLCHLFQKKDYVRTCIMDNGVSYRGCTVARTAGGLPCQAW 134
QY      50  ---LDAQSGLASAPVSGAGNHSCYCRNPDDPRGPGWCYVS-----85
Db      135  SRRFPNDHKYVTPPKNGL-EENFCRNPDDGDPGRGWCYTTNRSVRFQSCGIGIKTCREAVCVL 193
QY      86  -----GEAGVPEK-RPCE--DLRCPET 104
Db      194  CNGEDYRGEVDVTESGRECORWDLOHPS 222

RESULT 15
JS0600
t-plasminogen activator (EC 3.4.21.68) gamma precursor - common vampire bat
N:Alternate names: tissue plasminogen activator
C:Species: Desmodus rotundus (common vampire bat)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: JS0600
R:Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Don
Gene 105, 229-237, 1991
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des
A:Reference number: JS0597; MUID:92039036; PMID:1937019
A:Accession: JS0600
A:Molecule type: mRNA
A:Residues: 1-394 <KRA>
A:Cross-references: GB:M63990; NID:gl66078; PIDN:AAA1595.1; PID:gl66079
A>Note: the authors translated the codon ATC for residue 75 as Thr
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-36/Domain: propeptide #status predicted <PRO>
F:37-394/Product: plasminogen activator gamma #status predicted <PLA>
F:45-126/Domain: kringle homology <KR>
F:143-388/Domain: trypsin homology <TRY>
F:45-126,66-108,97-121,131-262,174-190,182-251,276-351,308-324,341-369/Disulfide bonds:
F:142-143/Cleavage site: His-Ser (plasmin) #status predicted
F:189,238,345/Active site: His, Asp, Ser #status predicted
F:315/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          9.5%; Score 132; DB 2; Length 394;
Best Local Similarity 32.7%; Pred. No. 0.0016;
Matches 33; Conservative 9; Mismatches 39; Indels 20; Gaps 5;

QY      17  AEAYGS--GGCFWDNGHLYREDOTSPAPGLRCLNW-----LDAQSGLASAPVSGAGNH 67
Db      35  SRAYGDPHATCYKDQGVYRGVTSBSSGAQCINWNSNLLIRTYNGRMPEAVKLGIGNH 94
QY      68  SYCRNPDDPRGPGWCYV-----SGEAGVPEKRPCEDLRC 101
Db      95  NYCRNPDGASK-PWCYVIKARKFTSBSVCP---VCSKATC 131
```

Search completed: November 25, 2003, 13:33:47  
Job time : 21 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:24:42 ; Search time 18 Seconds  
(without alignments)  
687.112 Million cell updates/sec

Title: US-10-057-951-2

Perfect score: 1393

Sequence: 1 MLLAWQAFVLSNMLAEAY.....PVDPEGSGPLMGQAGTPGA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	160.5	11.5	566	1	TPA_BOVIN
2	154.5	11.1	562	1	TPA_HUMAN
3	154	11.1	559	1	TPA_RAT
4	150	10.8	653	1	HGFA_MOUSE
5	147.5	10.6	559	1	TPA_MOUSE
6	146.5	10.5	655	1	HGFA_HUMAN
7	142	10.2	810	1	PLMN_HUMAN
8	138	9.9	169	1	PLMN_RAT
9	138	9.9	433	1	UROK_PAPCY
10	135.5	9.7	442	1	UROK_FIG
11	134	9.6	431	1	URTB_DESRO
12	134	9.6	477	1	URT2_DESRO
13	134	9.6	716	1	HGFL_MOUSE
14	132	9.5	394	1	URTG_DESRO
15	132	9.5	431	1	UROK_HUMAN
16	130.5	9.4	461	1	KRM2_MOUSE
17	130	9.3	810	1	PLMN_ERIEU
18	128.5	9.2	434	1	UROK_CHICK
19	127	9.1	433	1	UROK_BOVIN
20	126.5	9.1	728	1	HGF_MOUSE
21	124.5	8.9	810	1	PLMN_MACMU
22	124	8.9	622	1	THRB_HUMAN
23	124	8.9	728	1	HGF_HUMAN
24	124	8.9	812	1	PLMN_BOVIN
25	123	8.8	593	1	FA12_BOVIN
26	122.5	8.8	4548	1	APOA_HUMAN
27	122	8.8	603	1	FA12_CAVPO
28	121.5	8.7	462	1	KRM2_HUMAN
29	121.5	8.7	728	1	HGF_RAT
30	120.5	8.7	333	1	PLMN_CANFA
31	120	8.6	433	1	UROK_MOUSE
32	120	8.6	790	1	PLMN_PIG
33	119.5	8.6	477	1	URT1_DESRO

34	118.5	8.5	615	1	FA12_HUMAN
35	118	8.5	432	1	UROK_RAT
36	117.5	8.4	812	1	PLMN_MOUSE
37	117	8.4	625	1	THRB_BOVIN
38	116	8.3	1420	1	APOA_MACMU
39	115.5	8.3	711	1	HGFL_HUMAN
40	114	8.2	452	1	KRM1_XENLA
41	114	8.2	473	1	KRM1_MOUSE
42	114	8.2	473	1	KRM1_RAT
43	114	8.2	475	1	KRM1_HUMAN
44	114	8.2	618	1	THRB_MOUSE
45	110.5	7.9	685	1	ROR1_DROME

#### ALIGNMENTS

RESULT 1  
ID TPA\_BOVIN STANDARD; PRT; 566 AA.  
AC Q28198;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tpa)  
DE (t-PA) (t-plasminogen activator).  
GN PLAT.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RA Ravn P., Berglund L., Petersen T.E.;  
RT "Cloning and characterization of the bovine plasminogen activators uPA  
and tPA.";  
RL Int. Dairy J. 5:605-617(1995).  
CC -!- FUNCTION: CONVERTS THE ABUNDANT, BUT INACTIVE, ZYMAGEN PLASMINOGEN  
TO PLASMIN BY HYDROLYZING A SINGLE ARG-VAL BOND IN PLASMINOGEN. BY  
CONTROLLING PLASMIN-MEDIATED PROTEOLYSIS, IT PLAYS AN IMPORTANT  
ROLE IN TISSUE REMODELING AND DEGRADATION, IN CELL MIGRATION AND  
MANY OTHER PHYSIOLOGICAL EVENTS.  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
plasminogen to form plasmin.  
CC -!- SUBUNIT: HETERODIMER OF CHAIN A AND CHAIN B HELD BY A DISULFIDE  
BOND.  
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR.  
CC -!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER  
PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER  
ARG-314 CATALYZED BY PLASMIN. TISSUE KALLIKREIN OR FACTOR XA.  
CC -!- MISCELLANEOUS: BINDS TO THE KRINGLE STRUCTURE OF THE FIBRIN A  
CHAIN. BINDING TO FIBRIN ENHANCES ITS CATALYTIC ACTIVITY.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.  
CC -!- SIMILARITY: Contains 2 kringle domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X85800; CAA59795.1; --  
CC HSPSP; P00750; IRTF.  
CC MEROPS; S01.232; --  
CC InterPro; IPR001314; Chymotrypsin.  
CC InterPro; IPR006209; EGF like.  
CC InterPro; IPR000083; Fibrinctnl.

```

DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00039; fnl; 1.
DR Pfam; PF00051; kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PRINTS; PRO0018; KRINGLE.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00058; FNL; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SBR; 1.
KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Signal; Kringle; EGF-like domain; Repeat; Signal.
FT SIGNAL 1 21
FT PROPEP 22 33
FT CHAIN 34 566
FT CHAIN 34 314
FT CHAIN 315 566
FT DOMAIN 40 82
FT DOMAIN 83 121
FT DOMAIN 128 209
FT DOMAIN 219 300
FT DOMAIN 315 566
FT ACT_SITE 361 361
FT ACT_SITE 410 410
FT ACT_SITE 517 517
FT DISULFID 42 72
FT DISULFID 70 79
FT DISULFID 87 98
FT DISULFID 92 109
FT DISULFID 111 120
FT DISULFID 128 209
FT DISULFID 149 191
FT DISULFID 180 204
FT DISULFID 219 300
FT DISULFID 240 282
FT DISULFID 271 295
FT DISULFID 303 434
FT DISULFID 346 362
FT DISULFID 354 423
FT DISULFID 448 523
FT DISULFID 480 496
FT DISULFID 513 541
FT CARBOHYD 153 153
FT CARBOHYD 487 487
SQ SEQUENCE 566 AA; 63701 MW; 2EB5EBE4E32276C3 CRC64;

Query Match 11.5%; Score 160.5; DB 1; Length 566;
Best Local Similarity 35.5%; Pred. No. 5.9e-06;
Matches 38; Conservative 11; Mismatches 45; Indels 13; Gaps 4;

QY 16 LAEAYSGGCFWNGHLYREDQTSAPGLRCLNWLDAQSLASAPVS-----GAGN 66
DB 119 LCIDATATCYKQGVAYRGYRTWTAEISGACANW--NSSGLAMKPYSGRRPNAIRLGLGN 176
QY 67 HSYCRNPDEPRGVCVSGEAGVPEKRPCEDLRCPETTSQALPAFT 113
DB 177 HNYCRNPDDQSK-PWCYVF-KAGKYISEFCSTPACAKVAEDGDCYT 221

```

```

RESULT 2
TPA_HUMAN
ID TPA_HUMAN STANDARD; PRT: 562 AA.
AC P00750; Q15103;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)
DE (t-PA) (t-plasminogen activator) (Alteplase) (Retepase).
GN PLAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=83115262; PubMed=6337343;
RA Pennica D., Holmes W.E., Kohn W.J., Harkins R.N., Vohar G.A.,
RA Ward C.A., Bennett W.F., Yelverton E., Seeburg P.H., Heyneker H.L.,
RA Goeddel D.V., Collen D.;
RT "Cloning and expression of human tissue-type plasminogen activator
RT cDNA in E. coli.";
RL Nature 301:214-221(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Petal lung;
RX MEDLINE=88262579; PubMed=3133640;
RA Sasaki H., Saito Y., Hayashi M., Otsuka K., Niwa M.;
RT "Nucleotide sequence of the tissue-type plasminogen activator cDNA
RT from human fetal lung cells.";
RL Nucleic Acids Res. 16:5695-5695(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=88054470; PubMed=2824147;
RA Reddy V.B., Garramone A.J., Sasak H., Wei C.-M., Watkins P., Galli J.,
RA Hsiung N.;
RT "Expression of human uterine tissue-type plasminogen activator in
RT mouse cells using BPV vectors.";
RL DNA 6:461-472(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=86196143; PubMed=3009482;
RA Frieznar Degen S.J., Rajput B., Reich E.;
RT "The human tissue plasminogen activator gene.";
RL J. Biol. Chem. 261:6972-6985(1986).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=84298137; PubMed=6089198;
RA Ny T., Elgh F., Lund B.;
RT "The structure of the human tissue-type plasminogen activator gene:
RT correlation of intron and exon structures to functional and
RT structural domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359(1984).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=86284200; PubMed=3090401;
RA Harris T.J., Patel T., Marston F.A., Little S., Emtage J.S.,
RA Odenaker G., Volckaert G., Rembauts W., Billiau A., Somer P.;
RT "Cloning of cDNA coding for human tissue-type plasminogen activator
RT and its expression in Escherichia coli.";
RL Mol. Biol. Med. 3:279-292(1986).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC TISSUE=umbilical vein;
RX MEDLINE=90192129; PubMed=2107528;
RA Siebert P.D., Fong K.;
RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from
RT human endothelial cells.";
RL Nucleic Acids Res. 18:1086-1086(1990).
RN [8]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;

```

RA MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Villalón D.K., Murry D.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
RA Richards S., Worley K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [9]  
RP SEQUENCE OF 212-361 FROM N.A.  
RX MEDLINE=83169656; PubMed=6572897;  
RA Edlund T., Ny T., Raanby M., Heden L.-O., Palm G., Holmgren E.,  
RA Josephson S.;  
RT "Isolation of cDNA sequences coding for a part of human tissue  
RT plasminogen activator.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352 (1983).  
RN [10]  
RP SEQUENCE OF 1-36 FROM N.A.  
RX MEDLINE=85289338; PubMed=3161893;  
RA Fisher R., Waller E.K., Grossi G., Thompson D., Tizard R.,  
RA Schlemming W.-D.;  
RT "Isolation and characterization of the human tissue-type plasminogen  
RT activator structural gene including its 5' flanking region.";  
RL J. Biol. Chem. 260:11223-11230 (1985).  
RN [11]  
RP SEQUENCE OF 31-562 FROM N.A.  
RX MEDLINE=91291340; PubMed=1368681;  
RA Itagaki Y., Yasuda H., Morinaga T., Mitsuda S., Higashio K.;  
RT "Purification and characterization of tissue plasminogen activator  
RT secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";  
RL Agric. Biol. Chem. 55:1225-1232 (1991).  
RN [12]  
RP SEQUENCE OF 36-562.  
RC TISSUE-Melanoma;  
RX MEDLINE=85000468; PubMed=6433976;  
RA Pohl G., Kaellstroem M., Bergsdorf N., Wallen P., Joernvall H.;  
RT "Tissue plasminogen activator: peptide analyses confirm an indirectly  
RT derived amino acid sequence, identify the active site serine residue,  
RT establish glycosylation sites, and localize variant differences.";  
RL Biochemistry 23:3701-3707 (1984).  
RN [13]  
RP SEQUENCE OF 33-52 AND 311-330.  
RC TISSUE-Melanoma;  
RX MEDLINE=83209620; PubMed=6682760;  
RA Wallen P., Pohl G., Bergsdorf N., Raanby M., Ny T., Joernvall H.;  
RT "Purification and characterization of a melanoma cell plasminogen  
RT activator.";  
RL Eur. J. Biochem. 132:681-686 (1983).  
RN [14]  
RP STRUCTURE OF CARBOHYDRATES.  
RX MEDLINE=90092112; PubMed=2513186;  
RA Pfeiffer G., Schmidt M., Strube K.-H., Geyer R.;  
RT "Carbohydrate structure of recombinant human uterine tissue  
RT plasminogen activator expressed in mouse epithelial cells.";  
RL Eur. J. Biochem. 186:273-286 (1989).  
RN [15]  
RP CARBOHYDRATE-LINKAGE SITE THR-96.  
RX MEDLINE=91159408; PubMed=1900431;  
RA Harris R.J., Leonard C.K., Guzzetta A.W., Spellman M.W.;  
RT "Tissue plasminogen activator has an O-linked fucose attached to

RT threonine-61 in the epidermal growth factor domain.";  
RL Biochemistry 30:2311-2314 (1991).  
RN [16]  
RX MEDLINE=91244765; PubMed=1645136;  
RA Vlahos C.J., Wilhelm O.G., Hassell T., Jaskunas S.R., Bang N.U.;  
RT "Disulfide pairing of the recombinant kringle-2 domain of tissue  
RT plasminogen activator produced in *Escherichia coli*.";  
RL J. Biol. Chem. 266:10070-10072 (1991).  
RN [17]  
RX X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF CATALYTIC DOMAIN.  
RP MEDLINE=96200385; PubMed=8613982;  
RA Lamba D., Bauer M., Huber R., Fischer S., Rudolph R., Kohnert U.,  
RA Bode W.;  
RT "The 2.3 Å crystal structure of the catalytic domain of recombinant  
RT two-chain human tissue-type plasminogen activator.";  
RL J. Mol. Biol. 258:117-135 (1996).  
RN [18]  
RX X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF CATALYTIC DOMAIN.  
RP MEDLINE=97449126; PubMed=9305622;  
RA Renatus M., Engh R.A., Stubbs M.T., Huber R., Fischer S., Kohnert U.,  
RA Bode W.;  
RT "Lysine 156 promotes the anomalous proenzyme activity of tPA: X-ray  
RT crystal structure of single-chain human tPA.";  
RL EMBO J. 16:4797-4805 (1997).  
RN [19]  
RX X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF KRINGLE 2.  
RP MEDLINE=92118803; PubMed=1310033;  
RA de Vos A., Ultsch M.H., Kelley R.F., Padmanabhan K., Tulinsky A.,  
RA Westbrook M.L., Kosiakof A.A.;  
RT "Crystal structure of the kringle 2 domain of tissue plasminogen  
RT activator at 2.4-Å resolution.";  
RL Biochemistry 31:270-279 (1992).  
RN [20]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=90122799; PubMed=2558718;  
RA Byeon I.-J.L., Kelley R.F., Llinas M.;  
RT "1H NMR structural characterization of a recombinant kringle 2 domain  
RT from human tissue-type plasminogen activator.";  
RL Biochemistry 28:9350-9360 (1989).  
RN [21]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=91200042; PubMed=1901789;  
RA Byeon I.-J.L., Kelley R.F., Llinas M.;  
RT "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR  
RT assignments and secondary structure.";  
RL Eur. J. Biochem. 197:155-165 (1991).  
RN [22]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=92106329; PubMed=1762144;  
RA Byeon I.-J.L., Llinas M.;  
RT "Solution structure of the tissue-type plasminogen activator kringle  
RT 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic  
RT drug.";  
RL J. Mol. Biol. 222:1035-1051 (1991).  
RN [23]  

Query Match 11.1%; Score 154.5; DB 1; Length 562;  
Best Local Similarity 39.6%; Pred. No. 1.8e-05;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 25 CFWDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 75  
Db 127 CYEQGISYRGVTWSTAEGSACTNW--NSSALAKPYSGRRPDALRLGLGNHYCRNPD 184  
QY 76 DPRGPWCYVSGEAGVPEKRPCEDLRCPETTS 106  
Db 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

RESULT 3  
TPA\_RAT  
ID\_TPA\_RAT STANDARD; PRT; 559 AA.



AC DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)  
 DE (t-PA) (t-plasminogen activator).  
 GN PLAT.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89170114; PubMed=3148445;  
 RA Ny T., Leonardson G., Heueh A.J.W.;  
 RT "Cloning and characterization of a cDNA for rat tissue-type  
 RT plasminogen activator.";  
 RL DNA 7:671-677(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90130448; PubMed=2105315;  
 RA Feng P., Ohlsson M., Ny T.;  
 RT "The structure of the TATA-less rat tissue-type plasminogen activator  
 RT gene. Species-specific sequence divergences in the promoter predict  
 RT differences in regulation of gene expression.";  
 RL J. Biol. Chem. 265:2022-2027(1990).  
 CC -!- FUNCTION: CONVERTS THE ABUNDANT, BUT INACTIVE, ZMOGEN PLASMINOGEN  
 CC TO PLASMIN BY HYDROLYZING A SINGLE ARG-VAL BOND IN PLASMINOGEN. BY  
 CC CONTROLLING PLASMIN-MEDIATED PROTEOLYSIS, IT PLAYS AN IMPORTANT  
 CC ROLE IN TISSUE REMODELING AND DEGRADATION, IN CELL MIGRATION AND  
 CC MANY OTHER PHYSIOLOGICAL EVENTS.  
 CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
 CC plasminogen to form plasmin.  
 CC -!- SUBUNIT: HETERODIMER OF CHAIN A AND CHAIN B HELD BY A DISULFIDE  
 CC BOND.  
 CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR.  
 CC -!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER  
 CC PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER  
 CC ARG-308 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.  
 CC -!- MISCELLANEOUS: BINDS TO THE KRINGLE STRUCTURE OF THE FIBRIN A  
 CC CHAIN. BINDING TO FIBRIN ENHANCES ITS CATALYTIC ACTIVITY.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 fibronectin type I domain.  
 CC -!- SIMILARITY: Contains 2 kringle domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M23697; AAA41812.1; -;  
 CC EMBL; M31197; AAA42261.1; -;  
 CC EMBL; M31185; AAA42261.1; JOINED.  
 CC EMBL; M31186; AAA42261.1; JOINED.  
 CC EMBL; M31187; AAA42261.1; JOINED.  
 CC EMBL; M31188; AAA42261.1; JOINED.  
 CC EMBL; M31189; AAA42261.1; JOINED.  
 CC EMBL; M31190; AAA42261.1; JOINED.  
 CC EMBL; M31191; AAA42261.1; JOINED.  
 CC EMBL; M31192; AAA42261.1; JOINED.  
 CC EMBL; M31193; AAA42261.1; JOINED.  
 CC EMBL; M31194; AAA42261.1; JOINED.  
 CC EMBL; M31195; AAA42261.1; JOINED.  
 CC EMBL; M31196; AAA42261.1; JOINED.  
 CC EMBL; A19618; CAA01482.1; -;  
 CC PIR; A35029; A35029.  
 CC HSSP; P00750; 1RTF.  
 CC MEROPS; S01.232; -;  
 CC InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR006209; EGF-like  
 DR InterPro; IPR00083; Fibrinctnl.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR000001; Kringle.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00039; fnl; 1.  
 DR Pfam; PF00051; kringle; 2.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle; 2.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00058; FN1; 1.  
 DR SMART; SM00130; KR; 2.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 2.  
 DR PROSITE; PS00070; KRINGLE\_2; 2.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
 KW Plasma; Kringle; EGF-like domain; Repeat; Signal.  
 FT SIGNAL 1 17  
 FT PROPEP 18 29  
 FT CHAIN 30 559  
 FT CHAIN 30 308  
 FT CHAIN 309 559  
 FT DOMAIN 36 78  
 FT DOMAIN 79 117  
 FT DOMAIN 124 205  
 FT DOMAIN 213 294  
 FT DOMAIN 309 559  
 FT ACT\_SITE 355 355  
 FT ACT\_SITE 404 404  
 FT ACT\_SITE 510 510  
 FT DISULFID 38 68  
 FT DISULFID 66 75  
 FT DISULFID 83 94  
 FT DISULFID 88 105  
 FT DISULFID 107 116  
 FT DISULFID 124 205  
 FT DISULFID 145 187  
 FT DISULFID 176 200  
 FT DISULFID 213 294  
 FT DISULFID 234 276  
 FT DISULFID 265 289  
 FT DISULFID 297 428  
 FT DISULFID 340 356  
 FT DISULFID 348 417  
 FT DISULFID 442 516  
 FT DISULFID 474 490  
 FT DISULFID 506 534  
 FT CARBOHYD 149 149  
 FT CARBOHYD 481 481  
 FT CONFLICT 380 380  
 SQ SSQUENCE 559 AA; 62903 MW; 7DBD3809C1D1C921 CRC64;

Query Match 11.1%; Score 154; DB 1; length 559;  
 Best Local Similarity 32.4%; Pred. No. 2e-05;

Matches 45; Conservative 14; Mismatches 54; Indels 26; Gaps 6;

Qy 25 CFWDNGHLYREDQTSFAPGLRCLNWLDAGSLASAPVS-----GAGNHVYCRNPDE 75

Db 124 CFEGQGITYRGVTWSTENGAEICNW--NSSALQKPYSAARRPNAIKLGLGNHNYCRNPDR 181

Qy 76 DPRGPWCYVSGEAGVPEKPCEDLRPE-----TTSQALPAFTTEIGAESEGPG 124

```

Db 182 DVK-PWCYVF-KAGYTTBFCSTPACPKGPTDCVVGKGVYRGTHSFTT--SKASCLPW 237
Qy 125 ADEQVQFAPANALPARSE 143
Db 238 NSMILIGHTYTAWRANQA 256

RESULT 4
HGFA MOUSE
ID HGFA_MOUSE STANDARD; PRT; 653 AA.
AC Q9R098; O3JKV4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hepatocyte growth factor activator precursor (EC 3.4.21.-) (HGF
activator) (HGFA).
GN HGFA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Itoh H., Kataoka H., Koono H.;
RT "Mouse hepatocyte growth factor activator.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA van Aelsberg J.S., Sehgal S., Kukes A., Brady C., Barasch J.,
RA Yang J., Huan Y.;
RT "Activation of HGF by endogenous HGF activator is required for
metanephric kidney morphogenesis in vitro.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTIVATES HEPATOCYTE GROWTH FACTOR (HGF) BY CONVERTING
IT FROM A SINGLE CHAIN TO A HETERODIMERIC FORM (BY SIMILARITY).
CC -!- SUBUNIT: DIMER OF A SHORT CHAIN AND A LONG CHAIN LINKED BY A
DISULFIDE BOND (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED AS AN INACTIVE SINGLE-CHAIN
PRECURSOR AND IS THEN ACTIVATED TO A HETERODIMERIC FORM (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.
CC -!- SIMILARITY: Contains 1 fibronectin type II domain.
CC -!- SIMILARITY: Contains 1 kringle domain.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF099017; AAF02489.1; -.
CC EMBL; AF224724; AAF34712.1; -.
CC HSSP; P00763; 1DPO.
CC MEROPS; S01.228; -.
CC MGD; MGI:1859281; Hgfac.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibronctn1.
DR InterPro; IPR000582; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.

```

```

DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000995; FN_Type_II; 1.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF_2.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Glycoprotein; Plasma; Serine protease; Kringle; Signal;
KW EGF-like domain; Repeat; Zymogen.
FT SIGNAL 1 29
FT PROPEP 30 369
FT CHAIN 370 405
FT CHAIN 406 653
FT DOMAIN 105 145
FT DOMAIN 157 195
FT DOMAIN 197 237
FT DOMAIN 238 276
FT DOMAIN 283 364
FT DOMAIN 406 653
FT ACT SITE 445 445
FT ACT SITE 495 495
FT ACT SITE 596 596
FT DISULFID 105 130
FT DISULFID 119 145
FT DISULFID 161 172
FT DISULFID 166 183
FT DISULFID 185 194
FT DISULFID 199 227
FT DISULFID 225 234
FT DISULFID 242 253
FT DISULFID 247 264
FT DISULFID 266 275
FT DISULFID 283 364
FT DISULFID 304 346
FT DISULFID 335 359
FT DISULFID 392 519
FT DISULFID 430 446
FT DISULFID 438 508
FT DISULFID 533 602
FT DISULFID 565 581
FT DISULFID 592 620
FT CARBOHYD 39 39
FT CARBOHYD 47 47
FT CARBOHYD 63 63
FT CARBOHYD 287 287
FT CARBOHYD 466 466
FT CARBOHYD 544 544
FT CONFLICT 164 164
SQ SEQUENCE 653 AA; 70567 MW; 88B4B2055DF7FDC CRC64;

```

Query Match 10.8%; Score 150; DB 1; Length 653;

Best Local Similarity 32.6%; Pred No. 5, 2e-05;

Matches 46; Conservative 17; Mismatches 46; Indels 32; Gaps 7;

Qy 25 CFWDNCHLYREDQTSAPGLRCLNW-----LDAQSLASAPVSGAGNHSYCRNPDED 76

Db 283 CFLGNGTEYRGVASTAASGLSCLAWNSDLLYQLHVDVS-VAAAVLLGLGPHAYCRNPDKD 341

Qy 77 PRGPWCYVSGEAGVPEKPRCEDLRCPETTSQALPAFTTEIQEASEGPGADEVOVFAPANA 136

Db	342	ER-PWCYVVKDNLASWE-----YCRLFACESLARVHSQTPE-----ILA----	A 380
Qy	137	LPARSEAAAVQPVIGISQVR 157	
Db	381	LP--ESAPAVRTCKRHKR 399	
RESULT 5			
TPA_MOUSE			
ID	TPA_MOUSE	STANDARD;	PRT; 559 AA.
AC	P11214; Q91VP2;		
DT	01-JUL-1989 (Rel. 11, Created)		
DT	15-SEP-2003 (Rel. 42, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)		
DE	(t-PA) (t-plasminogen activator).		
GN	PLAT.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88087303; PubMed=2826484;		
RA	Rickles R.J., Darrow A.L., Strickland S.;		
RT	"Molecular cloning of complementary DNA to mouse tissue plasminogen		
RT	activator mRNA and its expression during P9 teratocarcinoma cell		
RT	differentiation.";		
RL	J. Biol. Chem. 263:1563-1569(1988).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Mammary gland;		
RX	MEDLINE=22388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Shapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,		
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,		
RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
CC	-!- FUNCTION: CONVERTS THE ABUNDANT, BUT INACTIVE, ZMOGEN PLASMINOGEN		
CC	TO PLASMIN BY HYDROLYZING A SINGLE ARG-VAL BOND IN PLASMINOGEN. BY		
CC	CONTROLLING PLASMIN-MEDIATED PROTEOLYSIS, IT PLAYS AN IMPORTANT		
CC	ROLE IN TISSUE REMODELING AND DEGRADATION, IN CELL MIGRATION AND		
CC	MANY OTHER PHYSIOPATHOLOGICAL EVENTS.		
CC	-!- CATALYTIC ACTIVITY: Specific cleavage of Arg- -Val bond in		
CC	plasminogen to form plasmin.		
CC	-!- SUBUNIT: HETERODIMER OF CHAIN A AND CHAIN B HELD BY A DISULFIDE		
CC	BOND.		
CC	-!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR.		
CC	-!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER		
CC	PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER		
CC	ARG-308 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.		
CC	-!- MISCELLANEOUS: BINDS TO THE KRINGLE STRUCTURE OF THE FIBRIN A		
CC	CHAIN. BINDING TO FIBRIN ENHANCES ITS CATALYTIC ACTIVITY.		
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.		
CC	-!- SIMILARITY: Contains 1 EGF-like domain.		
CC	-!- SIMILARITY: Contains 1 fibronectin type I domain.		
CC	-!- SIMILARITY: Contains 2 kringle domains.		

CC	EMBL; J03520; AAA40470.1; -.		
CC	EMBL; BC011256; AAH11256.1; -.		
DR	PIR; A29941; A29941.		
DR	HSSP; P00750; IASH.		
DR	MEROPS; S01.232; -.		
DR	MGI; MGI:97610; Plat.		
DR	InterPro; IPR001314; Chymotrypsin.		
DR	InterPro; IPR006209; EGF like.		
DR	InterPro; IPR000083; Fibrinctnl.		
DR	InterPro; IPR006210; IEGF.		
DR	InterPro; IPR000001; Kringle.		
DR	InterPro; IPR001254; Ser_protease_Try.		
DR	Pfam; PF00008; EGF; 1.		
DR	Pfam; PF00039; fnl; 1.		
DR	Pfam; PF00051; kringle; 2.		
DR	Pfam; PF00089; trypsin; 1.		
DR	PRINTS; PR00722; CHYMOTRYPSIN.		
DR	PRINTS; PR00018; KRINGLE.		
DR	ProDom; PD000395; Kringle; 2.		
DR	SMART; SM00181; EGF; 1.		
DR	SMART; SM00058; FN1; 1.		
DR	SMART; SM00130; KR; 2.		
DR	SMART; SM00020; Tryp_SPC; 1.		
DR	PROSITE; PS00022; EGF_1; 1.		
DR	PROSITE; PS01186; EGF_2; 1.		
DR	PROSITE; PS01253; FIBRONECTIN_1; 1.		
DR	PROSITE; PS00021; KRINGLE_1; 2.		
DR	PROSITE; PS00070; KRINGLE_2; 2.		
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.		
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.		
DR	PROSITE; PS00135; TRYPSIN_SER; 1.		
KW	Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;		
KW	Plasma; Kringle; EGF-like domain; Repeat; Signal.		
FT	SIGNAL	1	17
FT	PROPEP	18	29
FT	CHAIN	30	559
FT	CHAIN	30	308
FT	CHAIN	309	559
FT	DOMAIN	36	78
FT	DOMAIN	79	117
FT	DOMAIN	124	205
FT	DOMAIN	213	294
FT	DOMAIN	309	559
FT	ACT_SITE	355	355
FT	ACT_SITE	404	404
FT	ACT_SITE	510	510
FT	DISULFID	38	68
FT	DISULFID	66	75
FT	DISULFID	83	94
FT	DISULFID	88	105
FT	DISULFID	107	116
FT	DISULFID	124	205
FT	DISULFID	145	187
FT	DISULFID	176	200
FT	DISULFID	213	294
FT	DISULFID	234	276
FT	DISULFID	265	289
FT	DISULFID	297	428
FT	DISULFID	340	356
FT	DISULFID	348	417
FT	DISULFID	442	516
FT	DISULFID	474	490
FT	DISULFID	506	534

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)



```
FT DISULFID 594 622 BY SIMILARITY.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 492 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 644 644 R -> Q (IN REF. 2).
SQ SEQUENCE 655 AA; 70681 MW; 2CF72F1E1B862ED7 CRC64;

Query Match 10.5%; Score 146.5; DB 1; Length 655;
Best Local Similarity 36.9%; Pred. No. 0.0001;
Matches 41; Conservative 10; Mismatches 41; Indels 19; Gaps 6;

QY 25 CFWDNGHLYREDQTSAPGLRCLNW-----LDAQSGLASAPVSGAGNHGYCRNPD 76
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 CFLGNGTGYRGVASTSAGSLCLAWNSDLLYQELHVDV-VGAALLGLGPHAYCRNPDND 344

QY 77 PRQPCWYVSGEAGVP-----EKRCEDLRCPETTSQALPFTTETIQE-ASEG 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 ER-PWCYVVKDSALSWEYCRLEACESL----TRVQLSPDLLATLPBPASPG 390

RESULT 7
PLMN HUMAN STANDARD; PRT; 810 AA.
AC P00747;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Plasminogen precursor (EC 3.4.21.7) [Contains: Angiostatin].
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90202879; PubMed=2318848;
RA Petersen T.E., Martzen M.R., Ichinose A., Davie E.W.;
RT "Characterization of the gene for human plasminogen, a key proenzyme
RT in the fibrinolytic system.";
RL J. Biol. Chem. 265:6104-6111(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87162490; PubMed=3030813;
RA Forsgren M., Raden B., Israelsson M., Larsson K., Heden L.-O.;
RT "Molecular cloning and characterization of a full-length cDNA clone
RT for human plasminogen.";
RL FEBS Lett. 213:254-260(1987).
RN [3]
RP SEQUENCE OF 20-810.
RA Sottrup-Jensen L., Petersen T.E., Magnusson S.;
RL Submitted (JUL-1977) to the PIR data bank.
RN [4]
RP SEQUENCE OF 292-810 FROM N.A.
RX MEDLINE=85023311; PubMed=6148961;
RA Malinowski D.P., Sadler J.E., Davie E.W.;
RT "Characterization of a complementary deoxyribonucleic acid coding for
RT human and bovine plasminogen.";
RL Biochemistry 23:4243-4250(1984).
RN [5]
RP SEQUENCE OF 20-100.
RX MEDLINE=75093329; PubMed=122932;
RA Wiman B., Wallen P.;
RT "Structural relationship between 'glutamic acid' and 'lysine' forms
RT of human plasminogen and their interaction with the NH2-terminal
RT activation peptide as studied by affinity chromatography.";
RL Eur. J. Biochem. 50:489-494(1975).
RN [6]
RP SEQUENCE OF 95-580; 581-626; 657-700 AND 732-810.
RA Sottrup-Jensen L., Claeys H., Zajdel M., Petersen T.E., Magnusson S.;
RL (In) Davidson J.F., Rowan R.M., Samama M.M., Desnoyers P.C. (eds.);
RL Progress in chemical fibrinolysis and thrombolysis, pp.3:191-209,
```

```
RL Raven Press, New York (1978).
RN [7]
RP SEQUENCE OF 493-604.
RX MEDLINE=76043692; PubMed=126863;
RA Wiman B., Wallen P.;
RT "Amino-acid sequence of the cyanogen-bromide fragment from human
RT plasminogen that forms the linkage between the plasmin chains.";
RL Eur. J. Biochem. 58:539-547(1975).
RN [8]
RP SEQUENCE OF 581-810.
RX MEDLINE=77225245; PubMed=142009;
RA Wiman B.;
RT "Primary structure of the B-chain of human plasmin.";
RL Eur. J. Biochem. 76:129-137(1977).
RN [9]
RP ACTIVE SITE.
RX MEDLINE=73149248; PubMed=4694729;
RA Robbins K.C., Bernabe P., Arzadon L., Summaria L.;
RT "The primary structure of human plasminogen. II. The histidine loop
RT of human plasmin: light (B) chain active center histidine sequence.";
RL J. Biol. Chem. 248:1631-1633(1973).
RN [10]
RP ACTIVE SITE.
RX MEDLINE=69234739; PubMed=4240117;
RA Groskopf W.R., Summaria L., Robbins K.C.;
RT "Studies on the active center of human plasmin. Partial amino acid;
RT sequence of a peptide containing the active center serine residue.";
RL J. Biol. Chem. 244:3590-3597(1969).
RN [11]
RP OMEGA-AMINOCARBOXYLIC ACID-BINDING SITES.
RX MEDLINE=82213905; PubMed=6919539;
RA Trexler M., Vali Z., Patchy L.;
RT "Structure of the omega-aminocarboxylic acid-binding sites of human
RT plasminogen. Arginine 70 and aspartic acid 56 are essential for
RT binding of ligand by kringle 4.";
RL J. Biol. Chem. 257:7401-7406(1982).
RN [12]
RP FIBRIN AND OMEGA-AMINOCARBOXYLIC ACID BINDING SITES.
RX MEDLINE=85054794; PubMed=6094526;
RA Vali Z., Patchy L.;
RT "The fibrin-binding site of human plasminogen. Arginines 32 and 34
RT are essential for fibrin affinity of the kringle 1 domain.";
RL J. Biol. Chem. 259:13690-13694(1984).
RN [13]
RP PHOSPHORYLATION SITE SER-597.
RX MEDLINE=97345939; PubMed=9201958;
RA Wang H., Ptorek M., Bretthauer R.K., Castellino F.J.;
RT "Serine-578 is a major phosphorylation locus in human plasma
RT plasminogen.";
RL Biochemistry 36:8100-8106(1997).
RN [14]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=88185329; PubMed=3356193;
RA Marti T., Schaller J., Rickli E.E., Schmid K., Kamerling J.P.,
RA Garwig G.J., van Halbeek H., Vliegthart J.F.;
RT "The N- and O-linked carbohydrate chains of human, bovine and porcine
RT plasminogen. Species specificity in relation to sialylation and
RT fucosylation patterns.";
RL Eur. J. Biochem. 173:57-63(1988).
RN [15]
RP CARBOHYDRATE-LINKAGE SITE SER-268.
RX MEDLINE=97207306; PubMed=9054441;
RA Pirie-Shepherd S.R., Stevens R.D., Andon N.L., Enghild J.J.,
RA Pizzo S.V.;
RT "Evidence for a novel O-linked sialylated trisaccharide on Ser-248 of
RT human plasminogen 2.";
RL J. Biol. Chem. 272:7409-7411(1997).
RN [16]
RP CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE=95042728; PubMed=7525077;
RA O'Reilly M.S., Holmgren L., Shing Y., Chen C., Rosenthal R.A.,
RA Moses M., Lane W.S., Cao Y., Sage E.H., Folkman J.;
RT "Angiostatin: a novel angiogenesis inhibitor that mediates the
```

RT suppression of metastases by a Lewis lung carcinoma.";  
RL Cell 79:315-328(1994).  
RN [17]  
RP CHARACTERIZATION OF ANGIOSTATIN.  
RX MEDLINE=97238710; PubMed=9102221;  
RA Sim B.K., O'Reilly M.S., Liang H., Fortier A.H., He W., Madsen J.W.,  
RA Lapcevic R., Nacy C.A.;  
RT "A recombinant human angiostatin protein inhibits experimental primary  
RT and metastatic cancer.";  
RL Cancer Res. 57:1329-1334(1997).  
RN [18]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 374-461.  
RX MEDLINE=92031502; PubMed=1657148;  
RA Mulichak A.M., Tulinsky A., Ravichandran K.G.;  
RT "Crystal and molecular structure of human plasminogen kringle 4  
RT refined at 1.9-A resolution.";  
RL Biochemistry 30:10576-10588(1991).  
RN [19]  
RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF 374-461.  
RX MEDLINE=92031503; PubMed=1657149;  
RA Wu T.-P., Padmanabhan K., Tulinsky A., Mulichak A.M.;  
RT "The refined structure of the epsilon-aminocaproic acid complex of  
RT human plasminogen kringle 4.";  
RL Biochemistry 30:10589-10594(1991).  
RN [20]  
RP X-RAY CRYSTALLOGRAPHY (1.67 ANGSTROMS) OF 376-454.  
RX Stet B., Yamano A., Whitlow M., Teeter M.M.;  
RT "Structure of human plasminogen kringle 4 at 1.68 Angstrom and 277 K.  
RT A possible structural role of disordered residues.";  
RL Acta Crystallogr. D 53:169-178(1997).  
RN [21]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 102-181.  
RX Mathews I.I., Vanderhoff-Hanaver P., Castellino F.J., Tulinsky A.;  
RT "Crystal structures of the recombinant kringle 1 domain of human  
RT plasminogen in complexes with the ligands epsilon-aminocaproic acid  
RT and trans-4-(aminomethyl)cyclohexane-1-carboxylic Acid.";  
RL Biochemistry 35:2567-2576(1996).  
RN [22]  
RP X-RAY CRYSTALLOGRAPHY (1.66 ANGSTROMS) OF 480-563.  
RX MEDLINE=98198034; PubMed=9521645;  
RA Chang Y., Mochalkin I., McCance S.G., Cheng B., Tulinsky A.,  
RA Castellino F.J.;  
RT "Structure and ligand binding determinants of the recombinant kringle  
RT 5 domain of human plasminogen.";  
RL Biochemistry 37:3258-3271(1998).  
RN [23]  
RP STRUCTURE BY NMR OF 96-184.  
RX MEDLINE=94237157; PubMed=8181475;  
RA Rejante M.R., Llinas M.;  
RT "1H-NMR assignments and secondary structure of human plasminogen  
RT kringle 1.";  
RL Eur. J. Biochem. 221:927-937(1994).  
RN [24]  
RP STRUCTURE BY NMR OF 96-184.  
RX MEDLINE=94237158; PubMed=8181476;  
RA Rejante M.R., Llinas M.;  
RT "Solution structure of the epsilon-aminohexanoic acid complex of  
RT human plasminogen kringle 1.";  
RL Eur. J. Biochem. 221:939-949(1994).  
RN [25]  
RP STRUCTURE BY NMR OF 183-354.  
RX MEDLINE=96194156; PubMed=8652577;  
RA Soehndel S., Hu C.-K., Marti D., Affolter M., Schaller J., Llinas M.,  
RA Rickli E.E.;  
RT "Recombinant gene expression and 1H NMR characteristics of the  
RT kringle (2 + 3) supermodule: spectroscopic/functional individuality  
RT of plasminogen kringle domains.";  
RL Biochemistry 35:2357-2364(1996).  
RN [26]  
RP STRUCTURE BY NMR OF 374-461.  
RX MEDLINE=90219023; PubMed=2157850;  
RA Atkinson R.A., Williams R.J.P.;

RT "Solution structure of the kringle 4 domain from human plasminogen by  
RT 1H nuclear magnetic resonance spectroscopy and distance geometry.";  
RL J. Mol. Biol. 212:541-552(1996).  
RN [27]  
RP VARIANTS PHE-374 AND THR-620.  
Query Match 10.28; Score 142; DB 1; Length 810;  
Best Local Similarity 37.2%; Pred. No. 0.00031;  
Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4;  
QY 25 CFWDNGHLYREDQTSAPGLRCLNWLDA-----OSGLASAPVSGAGNHSYCRNPDEPRG 79  
DB 103 CKTGNGKNGYRGTSTKNGITCKQKWSSTSPHRPFSATHPEGL-EENYCRNPDNDPQG 161  
QY 80 PWCYVSGEAGVPEKR--PCEDLRCP 103  
DB 162 PWCYTTD----PEKRYDYCDILECEE 183  
RESULT 8  
PLMN RAT  
ID PLMN RAT STANDARD; PRT; 169 AA.  
AC Q01177;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Plasminogen (EC 3.4.21.7) (Fragment).  
GN PLG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=91250378; PubMed=1645711;  
RA Kanalas J.J., Makker S.P.;  
RT "Identification of the rat Heymann nephritis autoantigen (GP330) as a  
RT receptor site for plasminogen.";  
RL J. Biol. Chem. 266:10825-10829(1991).  
CC -!- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS  
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING  
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION  
CC AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE  
CC GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN  
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH  
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,  
CC LAMININ AND VON WILLEBRAND FACTOR.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Lys-|-Xaa > Arg-|-Xaa;  
CC higher selectivity than trypsin. Converts fibrin into soluble  
CC products.  
CC -!- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN  
CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO  
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.  
CC -!- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN  
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMINOGEN SUBFAMILY.  
CC -!- SIMILARITY: Contains 5 kringle domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M62832; AAA41884.1; -.  
DR PIR; A40522; A40522.  
DR HSP; P00747; 1PMK.  
DR MEROPS; S01.233; -.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR003966; Prothrombin.

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----

EMBL; X51935; CAA36200.1; -  
PIR; S14687; UKBAY.  
HSSP; P00749; 1LMW.  
MEROPS; S01.231; -  
InterPro; IPR001314; Chymotrypsin.  
InterPro; IPR006209; EGF\_like.  
InterPro; IPR006210; IEGF.  
InterPro; IPR000001; Kringle.  
InterPro; IPR001254; Ser\_protease\_Try.  
Pfam; PF00051; kringle; 1.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PR00722; CHYMOTRYPSIN.  
PRINTS; PR00018; KRINGLE.  
ProDom; PD000395; Kringle; 1.  
SMART; SM00181; EGF; 1.  
SMART; SM00130; KR; 1.  
SMART; SM00020; Tryp\_SPC; 1.  
PROSITE; PS00022; EGF\_1; 1.  
PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
PROSITE; PS00021; KRINGLE\_1; 1.  
PROSITE; PS00070; KRINGLE\_2; 1.  
PROSITE; PS0240; TRYPSIN\_DOM; 1.  
PROSITE; PS00134; TRYPSIN\_HIS; 1.  
PROSITE; PS00135; TRYPSIN\_SER; 1.  
PROSITE; PS00136; TRYPSIN\_SER; 1.  
Kw; Plasminogen activation; Hydrolase; Serine protease; Glycoprotein; Kringle; EGF-like domain; Zymogen; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 433 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.  
FT CHAIN 21 176 CHAIN A (BY SIMILARITY).  
FT CHAIN 155 176 SHORT A CHAIN (A1) (BY SIMILARITY).  
FT CHAIN 178 433 CHAIN B (BY SIMILARITY).  
FT DOMAIN 26 62 EGF-LIKE.  
FT DOMAIN 69 150 KRINGLE.  
FT DOMAIN 151 177 CONNECTING PEPTIDE.  
FT DOMAIN 178 433 SERINE PROTEASE.  
FT DISULFID 30 38 BY SIMILARITY.  
FT DISULFID 32 50 BY SIMILARITY.  
FT DISULFID 52 61 BY SIMILARITY.  
FT DISULFID 167 298 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 208 224 BY SIMILARITY.  
FT DISULFID 216 287 BY SIMILARITY.  
FT DISULFID 315 384 BY SIMILARITY.  
FT DISULFID 347 363 BY SIMILARITY.  
FT DISULFID 374 402 BY SIMILARITY.  
FT ACT\_SITE 223 223 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 274 274 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 378 378 CHARGE RELAY SYSTEM.  
FT CARBOHYD 324 324 N-LINKED (GLCNAC... ) (BY SIMILARITY).  
SQ SEQUENCE 433 AA; 48595 MW; 816D22DFEDDC8792 CRC64;

Query Match 9.9%; Score 138; DB 1; Length 433;  
Best Local Similarity 32.7%; Pred. No. 0.00033;  
Matches 32; Conservative 14; Mismatches 36; Indels 16; Gaps 4

QY 25 CFWDNGHLYREDOTSPAPGRLCNWLDA-----QSGLASAPVSGAGNHSCYCRNPDEDP 77  
DB 69 CYEGNGHYRGKASTDTMGSRCLAWSATVLQCTYHAHRSDALQLGLGKHNYCRNP-NR 127  
QY 78 RGPWCYVSGAGVPEK-----RPECRLCPETTSOAL 109  
DB 128 RRPWCYV--QVGLKQRVQECMVHNCADGKPKSPPEEL 163

RESULT 10  
UROK\_PAPCY STANDARD; PRT; 442 AA.  
ID UROK\_PIG  
AC P04185;  
DT 20-MAR-1987 (Rel. 04, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)



28-FEB-2003 (Rel. 41, Last annotation update)  
Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)  
(U-plasminogen activator).  
PLAU.  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1] SEQUENCE FROM N.A.  
RP TISSUE=Kidney;  
RC MEDLINE=85087954; PubMed=6096832;  
RA Nagamine Y., Pearson D., Altus M.S., Reich E.;  
RX "cDNA and gene nucleotide sequence of porcine plasminogen activator.";  
RT Nucleic Acids Res. 12:9525-9541(1984).  
RL [2]  
RN REVISION TO 241.  
RP Nagamine Y.;  
RL Submitted (DEC-1986) to the PIR data bank.  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL; X01648; CAA25806.1; -;  
DR EMBL; X02724; CAA26511.1; -;  
DR PIR; A00932; UKFG.  
DR HSP; P00749; 1KDU.  
DR MEROPS; S01.231; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Ser.protease\_Try.  
DR Pfam; PF00051; kringle\_1.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00020; TRYD\_SPC; 1.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00021; KRINGLE\_2; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
DR PROSITE; PS00134; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Plasminogen activation; Hydrolyase; Serine protease; Glycoprotein;  
KW Kringle; EGF-like domain; Zymogen; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 442 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.  
FT CHAIN 21 188 CHAIN A (BY SIMILARITY).  
FT CHAIN 190 442 CHAIN B (BY SIMILARITY).  
FT DOMAIN 29 65 EGF-LIKE.  
FT DOMAIN 72 153 KRINGLE.  
FT DOMAIN 154 189 CONNECTING PEPTIDE.  
FT DOMAIN 190 442 SERINE PROTEASE.  
FT CARBOXYD 152 152 N-LINKED (GLCNAC. . .).  
FT DISULFID 33 41 BY SIMILARITY.  
FT DISULFID 35 53 BY SIMILARITY.  
FT DISULFID 55 64 BY SIMILARITY.  
FT DISULFID 179 310 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 220 236 BY SIMILARITY.  
FT DISULFID 228 299 BY SIMILARITY.

FT DISULFID 324 393 BY SIMILARITY.  
FT DISULFID 356 411 BY SIMILARITY.  
FT DISULFID 383 472 BY SIMILARITY.  
FT ACT\_SITE 235 235 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 286 286 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 387 387 CHARGE RELAY SYSTEM.  
FT CONFLICT 241 241 Q -> H (IN REF. 1; CAA25806).  
FT CONFLICT 242 242 Q -> H (IN REF. 1; CAA25811).  
FT CONFLICT 288 288 A -> GS (IN REF. 1; CAA25806).  
SQ SEQUENCE 442 AA; 49116 MW; EE32CFEF501321EE CRC64;  
Query Match 9.7%; Score 135.5; DB 1; Length 442;  
Best Local Similarity 36.9%; Pred. No. 0.00055;  
Matches 38; Conservative 7; Mismatches 37; Indels 21; Gaps 5;  
QY 25 CFWDNGHLYREDQTSAPAGRLCLNWLDAQSLG-----ASAPVS---GAGNHGYCRNPDED 77  
Db 72 CFEGNGHSYRGKANTNTGRCPLPNSATVLLNTYHAHRPDALQLGLGKHNYCRNP-NQ 130  
QY 78 RGPWCYVS-----GEAGVP-----EKRPCEDLRCPETTSQ 107  
Db 131 RRPWCYVQVGLKQLVQECMVPNCSGSGESHRPAYDKGNPFSTPE 173

## RESULT 11

URTB\_DESRO  
ID URTB\_DESRO STANDARD; PRT; 431 AA.  
AC P98121;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary plasminogen activator beta precursor (EC 3.4.21.68) (DSFA beta).  
OS Desmodus rotundus (Vampire bat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  
OC Desmodontinae; Desmodus.  
OX NCBI\_TaxID=9430;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Salivary gland;  
RX MEDLINE=92039036; PubMed=1937019;  
RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,  
RA Alagon A., Donner P., Schleuning W.D.;  
RT "The plasminogen activator family from the salivary gland of the  
RT vampire bat Desmodus rotundus: cloning and expression.";  
RL Gene 105:229-237(1991).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=93393059; PubMed=1309059;  
RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,  
RA Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,  
RA Donner P.;  
RT "Plasminogen activators from the saliva of Desmodus rotundus (common  
RT vampire bat): unique fibrin specificity.";  
RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
CC -!- FUNCTION: PROBABLY ESSENTIAL TO SUPPORT THE FEEDING HABITS OF THIS  
CC EXCLUSIVELY HAEMATOPHAGOUS ANIMAL. PROBABLE POTENT THROMBOLYTIC  
CC AGENT.  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -!- SUBUNIT: Monomer.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC -----
CC EMBL: M63989; AAA31594.1; -
CC PIR: J05099; J05099.
CC HSSP: P98119; 1A51.
CC MEROPS: S01.239; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR006209; EGF-like.
CC InterPro: IPR006210; IEGF.
CC InterPro: IPR000001; Kringle.
CC InterPro: IPR001254; Ser_protease_Try.
CC Pfam: PF00008; EGF_1.
CC Pfam: PF00051; kringle; 1.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PRINTS: PR00018; KRINGLE.
CC ProDom: PD000395; Kringle; 1.
CC SMART: SM00181; EGF; 1.
CC SMART: SM00130; KR; 1.
CC PROSITE: PS00022; TRYP_SPE; 1.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS00021; KRINGLE_1; 1.
CC PROSITE: PS50070; KRINGLE_2; 1.
CC PROSITE: PS50240; TRYPsin_DOM; 1.
CC PROSITE: PS00134; TRYPsin_HIS; 1.
CC PROSITE: PS00135; TRYPsin_SER; 1.
CC KMW: Plasmnogen activation; Hydrolase; Serine protease; Glycoprotein;
CC Kringle; EGF-like domain; Signal; Multigene family.
CC SIGNAL 1 36
CC FT CHAIN 37 431 SALIVARY PLASMINOGEN ACTIVATOR BETA.
CC FT DOMAIN 37 75 EGF-LIKE.
CC FT DOMAIN 82 163 KRINGLE.
CC FT DOMAIN 179 431 SERINE_PROTEASE.
CC FT ACT_SITE 226 431 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 275 275 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 382 382 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 41 52 BY SIMILARITY.
CC FT DISULFID 46 63 BY SIMILARITY.
CC FT DISULFID 65 74 BY SIMILARITY.
CC FT DISULFID 82 163 BY SIMILARITY.
CC FT DISULFID 103 145 BY SIMILARITY.
CC FT DISULFID 134 158 BY SIMILARITY.
CC FT DISULFID 168 299 BY SIMILARITY.
CC FT DISULFID 211 227 BY SIMILARITY.
CC FT DISULFID 219 288 BY SIMILARITY.
CC FT DISULFID 313 388 BY SIMILARITY.
CC FT DISULFID 345 361 BY SIMILARITY.
CC FT DISULFID 378 406 BY SIMILARITY.
CC FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 431 AA; 48221 MW; 699B5675B162CBF CRC64;

Query Match 9.6%; Score 134; DB 1; Length 431;
Best Local Similarity 38.6%; Pred. No. 0.00071;
Matches 27; Conservative 9; Mismatches 20; Indels 14; Gaps 3;

```

```

DE alpha-2) (BAT-PA) (T-plasminogen activator).
OS Desmodus rotundus (Vampire bat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;
OC Desmodontinae; Desmodus.
OX NCBI_TaxID=9430;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RX MEDLINE=92039036; PubMed=1937019;
RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,
RA Alagon A., Donner P., Schleuning W.D.;
RT "The plasminogen activator family from the salivary gland of the
RL vampire bat Desmodus rotundus: cloning and expression.";
RN Gene 105:229-237(1991).
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Salivary gland;
RX MEDLINE=9003687; PubMed=2509450;
RA Gardell S.J., Duong L.T., Diehl R.E., York J.D., Hare T.R.,
RA Register R.B., Jacobs J.W., Dixon R.A.F., Friedman P.A.;
RT "Isolation, characterization, and cDNA cloning of a vampire bat
RL salivary plasminogen activator.";
RN J. Biol. Chem. 264:17947-17952(1989).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=93393059; PubMed=1309059;
RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,
RA Kraetzschmar J., Haendler B., Langer G., Baldus B., Wilt W.,
RA Donner P.;
RT "Plasminogen activators from the saliva of Desmodus rotundus (Common
RT vampire bat): unique fibrin specificity.";
RL Ann. N.Y. Acad. Sci. 667:395-403(1992).
CC -1- FUNCTION: PROBABLY ESSENTIAL TO SUPPORT THE FEEDING HABITS OF THIS
CC EXCLUSIVELY HAEMATOPHAGOUS ANIMAL. PROBABLE POTENT THROMBOLYTIC
CC AGENT.
CC -1- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-val bond in
CC plasminogen to form plasmin.
CC -1- ENZYME REGULATION: ACTIVITY TOWARD PLASMINOGEN IS STIMULATED IN
CC THE PRESENCE OF FIBRIN I.
CC -1- SUBUNIT: Monomer.
CC -1- DOMAIN: THE FIBRINOCTIN TYPE-I DOMAIN MEDIATES BINDING TO FIBRIN,
CC AND THE KRINGLE DOMAIN APPARENTLY MEDIATES FIBRIN-INDUCED
CC STIMULATION OF ACTIVITY.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 fibronectin type I domain.
CC -1- SIMILARITY: Contains 1 kringle domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M63988; AAA31593.1; -
CC PIR: J05082; AAA31596.1; -
CC HSSP: P98119; 1A51.
CC MEROPS: S01.232; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR006209; EGF-like.
CC InterPro: IPR000083; Fibrinctn.
CC InterPro: IPR006210; IEGF.
CC InterPro: IPR000001; Kringle.
CC InterPro: IPR001254; Ser_protease_Try.
CC Pfam: PF00008; EGF; 1.
CC Pfam: PF00039; fn1; 1.
CC Pfam: PF00051; kringle; 1.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.

```



FT DISULFID 292 370 BY SIMILARITY.  
 FT DISULFID 313 352 BY SIMILARITY.  
 FT DISULFID 341 364 BY SIMILARITY.  
 FT DISULFID 379 457 BY SIMILARITY.  
 FT DISULFID 400 440 BY SIMILARITY.  
 FT DISULFID 428 452 BY SIMILARITY.  
 FT DISULFID 477 593 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 512 528 BY SIMILARITY.  
 FT DISULFID 607 672 BY SIMILARITY.  
 FT DISULFID 637 651 BY SIMILARITY.  
 FT DISULFID 662 690 BY SIMILARITY.  
 FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 19 19 P -> Q (IN GENOMIC SEQUENCE).  
 SQ SEQUENCE 716 AA; 80588 MW; BRCB02EF85213ACC CRC64;

Query Match 9.6%; Score 134; DB 1; Length 716;  
 Best Local Similarity 28.2%; Pred. No. 0.0012;  
 Matches 42; Conservative 9; Mismatches 46; Indels 52; Gaps 7;

QY 2 LIAAVQAFVSNMLAEAYSG-----GCFWDMGHLYREDQTSAPDQLCLMW 49  
 DB 80 LIPWTO-----HSHTOLYHSLCHLPQKDYVRTCIMDGVSYRGTVARTAGLPCQAM 134  
 QY 50 ---LDAGSGLASAPVSGAGNHSYCRNDEDPGRCMVCVS----- 85  
 DB 135 SRREPNDHKYPTPTKNGL-EEFCRNPDGPDRGMCYTTTKRSVRFQSGCIKTCEAVCVL 193  
 QY 86 -----GEAGVPEK-PPCE--DURCPET 104  
 DB 194 CNGEDYRGVDVTESGRCQWMDLQHHS 222

RESULT 14  
 URTG DESRO STANDARD; PRT; 394 AA.

AC P49150;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Salivary plasminogen activator gamma precursor (BC 3.4.21.68) (DSFA gamma).  
 DE Desmodus rotundus (Vampire bat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  
 OC Desmodontinae; Desmodus.  
 OC NCBI\_Taxid=9430;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RP TISSUE=Salivary gland;  
 RX MEDLINE=92039036; PubMed=1937019;  
 RA Krietzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,  
 RA Alagon A., Donner P., Schleuning W.D., Schleuning W.D.,  
 RT "The plasminogen activator family from the salivary gland of the  
 RT vampire bat Desmodus rotundus: cloning and expression.";  
 RL Gene 105:229-237(1991).  
 RN [2]  
 RP CHARACTERIZATION.  
 RP MEDLINE=9339059; PubMed=1309059;  
 RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,  
 RA Krietzschmar J., Haendler B., Langer G., Baldus B., Witt W.,  
 RA Donner P.;  
 RT "Plasminogen activators from the saliva of Desmodus rotundus (common  
 RT vampire bat): unique fibrin specificity.";  
 RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
 CC -!- FUNCTION: PROBABLY ESSENTIAL TO SUPPORT THE FEEDING HABITS OF THIS  
 CC EXCLUSIVELY HAEMATOPHAGOUS ANIMAL. PROBABLE POTENT THROMBOLYTIC  
 CC AGENT.  
 CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
 CC plasminogen to form plasmin.  
 CC -!- SUBUNIT: Monomer.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -!- SIMILARITY: Contains 1 kringle domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M63990; AAA31595.1; -  
 DR PIR; J50600; J50600.  
 DR HSSP; P98119; 1A51.  
 DR MEROPS; S01.239; -  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR000001; Kringle.  
 DR Pfam; PF00051; Kringle\_1.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle; 1.  
 DR SMART; SM00130; KR; 1.  
 DR SMART; SM00020; TRY-sec; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00070; KRINGLE\_2; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR PLASMINOGEN activation; Hydrolase; Serine protease; Glycoprotein;  
 KW Kringle; Signal; Multigene family.  
 FT SIGNAL 1 36  
 FT CHAIN 37 394  
 FT DOMAIN 45 126  
 FT ACT\_SITE 142 394  
 FT ACT\_SITE 189 189  
 FT ACT\_SITE 238 238  
 FT ACT\_SITE 345 345  
 FT DISULFID 45 126  
 FT DISULFID 66 108  
 FT DISULFID 97 121  
 FT DISULFID 131 262  
 FT DISULFID 174 190  
 FT DISULFID 182 251  
 FT DISULFID 276 351  
 FT DISULFID 308 324  
 FT DISULFID 341 369  
 FT CARBOHYD 315 315  
 SQ SEQUENCE 394 AA; 44105 MW; 9CCD6523D81FCD CRC64;

Query Match 9.5%; Score 132; DB 1; Length 394;  
 Best Local Similarity 32.7%; Pred. No. 0.00095;  
 Matches 33; Conservative 9; Mismatches 39; Indels 20; Gaps 5;

QY 17 AAAYGS--GCFWDMGHLYREDQTSAPDQLCLMW-----LDAGSGLASAPVSGAGNH 67  
 DB 35 SRAVYDPHATCYKQDGVYTRGTWSTSSGACCMNNSMLIRTYNGRMPEAVKLGLGNH 94  
 QY 68 SYCRNPDEDPGRCMVCVS-----SGEAGVPEKPPCEDLRC 101  
 DB 95 NYCRRNPDSASK-PMCYVTKARKFTSESCSVF---VCSKATC 131  
 RESULT 15  
 UROK HUMAN STANDARD; PRT; 431 AA.  
 AC P00749; O15844; Q16618; O969W6;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Urokinase-type plasminogen activator precursor (BC 3.4.21.73) (uPA)  
 DE (U-plasminogen activator).

GN PLAU.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=85215647; PubMed=2987867;  
 RA Riccio A., Grimaldi G., Verde P., Sebastio G., Boast S., Blasi F.;  
 RT "The human urokinase-plasminogen activator gene and its promoter";  
 RL Nucleic Acids Res. 13:2759-2771(1985).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RA Holmes W.E., Pennica D., Blaber M., Rey M.W., Guenzler W.A.,  
 RA Steffens G.J., Heyneker H.L.;  
 RT "Cloning and expression of the gene for pro-urokinase in *Escherichia coli*";  
 RL Biotechnology 3:923-929(1985).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=86056954; PubMed=2415429;  
 RA Nagai M., Hiramatsu R., Kaneda T., Hayasuke N., Arimura H.,  
 RA Nishida M., Suyama T.;  
 RT "Molecular cloning of cDNA coding for human prepro-urokinase";  
 RL Gene 36:183-188(1985).  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=85203359; PubMed=3888571;  
 RA Jacobs P., Cravador A., Lorlau R., Brockly F., Colau B., Chuchana P.,  
 RA van Elsen A., Herzog A., Bollen A.;  
 RT "Molecular cloning, sequencing, and expression in *Escherichia coli* of human prepro-urokinase cDNA";  
 RL DNA 4:139-146(1985).  
 RN [5]  
 RN SEQUENCE FROM N.A.  
 RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,  
 RA Nickerson D.A.;  
 RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RN SEQUENCE FROM N.A.  
 RC TISUB=Lung;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Casavant T.L., Scheetz T.E.,  
 RA Stabileton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Mullaly S.J.,  
 RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Bouffard G.G.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodighiero S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywicki M.I., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Mair M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RN SEQUENCE OF 66-431 FROM N.A.  
 RX MEDLINE=84272706; PubMed=6589620;  
 RA Verde P., Stoppelli M.P., Galetti P., di Nocera P., Blasi F.;  
 RT "Identification and primary sequence of an unspliced human urokinase RT poly(A)+ RNA";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:4727-4731(1984).  
 RN [8]  
 RN SEQUENCE OF 21-177,  
 RX MEDLINE=83055084; PubMed=6754569;  
 RA Gunzler W.A., Steffens G.J., Otting F., Kim S.-M.A., Frankus E.,

RA Flohe L.;  
 RT "The primary structure of high molecular mass urokinase from human urine. The complete amino acid sequence of the A chain";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 363:1155-1165(1982).  
 RN [9]  
 RN SEQUENCE OF 156-176 AND 179-224,  
 RX MEDLINE=83003608; PubMed=6749491;  
 RA Schaller J., Nick H., Rickli E.E., Gillesen D., Lergier W.,  
 RA Studer R.O.;  
 RT "Human low-molecular-weight urinary urokinase. Partial characterization and preliminary sequence data of the two polypeptide RT chains";  
 RL Eur. J. Biochem. 125:251-257(1982).  
 RN [10]  
 RN SEQUENCE OF 158-410,  
 RX MEDLINE=83055099; PubMed=6754572;  
 RA Steffens G.J., Gunzler W.A., Otting F., Frankus E., Flohe L.;  
 RT "The complete amino acid sequence of low molecular mass urokinase from human urine";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 363:1043-1058(1982).  
 RN [11]  
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE=9600858; PubMed=8591045;  
 RA Spraggon G., Phillips C., Nowak U.K., Ponting C.P., Saunders D.,  
 RA Dobson C.M., Stuart D.I., Jones E.Y.;  
 RT "The crystal structure of the catalytic domain of human urokinase-type plasminogen activator";  
 RL Structure 3:681-691(1995).  
 RN [12]  
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 159-411.  
 RX MEDLINE=20266327; PubMed=10805774;  
 RA Sperl S., Jacob U., Arroyo de Prada N., Sturzbecher J., Wilhelm O.G.,  
 RA Bode W., Magdolen V., Huber R., Moroder L.;  
 RT "4-aminoethylphenylguanidine derivatives as nonpeptidic highly selective inhibitors of human urokinase";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5113-5118(2000).  
 RN [13]  
 RN STRUCTURE BY NMR.  
 RX MEDLINE=89127526; PubMed=2536903;  
 RA Oswald R.E., Bogusky M.J., Bamberger M., Smith R.A.G., Dobson C.M.;  
 RT "Dynamics of the multidomain fibrinolytic protein urokinase from two-dimensional NMR";  
 RL Nature 337:579-582(1989).  
 RN [14]  
 RN STRUCTURE BY NMR OF 67-155.  
 RX MEDLINE=93003110; PubMed=1327118;  
 RA Li X., Smith R.A.G., Dobson C.M.;  
 RT "Sequential 1H NMR assignments and secondary structure of the kringe domain from urokinase";  
 RL Biochemistry 31:9562-9571(1992).  
 RN [15]  
 RN STRUCTURE BY NMR OF 67-155.  
 RX MEDLINE=94149701; PubMed=8107091;  
 RA Li X., Bokman A.M., Llinas M., Smith R.A.G., Dobson C.M.;  
 RT "Solution structure of the kringe domain from urokinase-type RT plasminogen activator";  
 RL J. Mol. Biol. 235:1548-1559(1994).  
 RN [16]  
 RN VARIANT LEU-141.  
 RX MEDLINE=96186279; PubMed=8652631;  
 RA Yoshimoto M., Ushiyama Y., Sakai M., Tamaki S., Hara H., Takahashi K.,  
 RA Sawasaki Y., Hanada K.;  
 RT "Characterization of single chain urokinase-type plasminogen activator with a novel amino-acid substitution in the kringe RT structure";  
 RL Biochim. Biophys. Acta 1293:83-89(1996).  
 RN [17]  
 RN VARIANT LEU-141.  
 RX MEDLINE=97218551; PubMed=9065988;  
 RA Conne B., Berczy M., Belin D.;  
 RT "Detection of polymorphisms in the human urokinase-type plasminogen activator gene";  
 RL Thromb. Haemost. 77:434-435(1997).

```
RN [18]
RP ERRATUM.
RA Come B., Berczy M., Belin D.;
RL Thromb. Haemost. 78:973-973(1997).
RN [19]
RP VARIANT LEU-141.
RX MEDLINE=97337920; PubMed=9194591;
RA Turkmen B., Schmitt M., Schmalfeldt B., Trommler P., Hell W.,
RA Creutzburg S., Graeff H., Magdolen V.;
RT "Mutational analysis of the genes encoding urokinase-type plasminogen
activator (uPA) and its inhibitor PAI-1 in advanced ovarian cancer.";
RL Electrophoresis 18:686-689(1997).
CC -1- FUNCTION: POTENT PLASMINOGEN ACTIVATOR AND IS CLINICALLY USED FOR
CC THERAPY OF THROMBOLYTIC DISORDERS.
CC -1- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
CC plasminogen to form plasmin.
CC -1- SUBUNIT: FOUND IN HIGH AND LOW MOLECULAR MASS FORMS. EACH CONSISTS
CC OF TWO CHAINS, A AND B. THE HIGH MOLECULAR MASS FORM CONTAINS A
CC LONG CHAIN A. CLEAVAGE OCCURS AFTER RESIDUE 155 IN THE LOW
CC MOLECULAR MASS FORM TO YIELD A SHORT A1 CHAIN.
CC -1- PHARMACEUTICAL: Available under the name Abbotkinase (Abbott). Used
CC in Pulmonary Embolism (PE) to initiate fibrinolysis.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02419; CAA26268.1; -
DR EMBL; M15476; AAA61253.1; -
DR EMBL; D00244; BAA00175.1; -
DR EMBL; D11143; BAA01919.1; -
DR EMBL; X02760; CAA26535.1; -
DR EMBL; AF377330; AAK53822.1; -
DR EMBL; BC013575; AAH13575.1; -
DR EMBL; K03226; AAC97138.1; -
DR EMBL; K02286; AAA61252.1; -
DR EMBL; A21571; CAA01559.1; -
DR EMBL; A18397; CAA01390.1; -
DR PIR; A00931; UKHU.
DR PDB; 1XDU; 31-OCT-93.

Query March 9.5%; Score 132; DB 1; Length 431;
Best Local Similarity 32.7%; Pred. No. 0.001;
Matches 33; Conservative 13; Mismatches 33; Indels 22; Gaps 5;

QY 25 CFWDNGHLVREDQTPAPGLRCINWLD-----QSGLASAPVSGAGNHSYCRNPDEDP 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 70 CYENGNGHFYRGKASTDITMGKPCLPWNSATVLOQTYHARSDALQLGKHNVCNRPD-NR 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 78 RGPWCYVSGEAGVPEKRP-----CEDLRCPETISQAL 109
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 129 RRPWCYV--QVGL--KPLVQECMVHDCADGKRPSPPEEL 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: November 25, 2003, 13:33:14  
Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 12:21:47 / Search time 40 Seconds  
(without alignments)  
1696.696 Million cell updates/sec

Title: US-10-057-951-2

Sequence: 1393  
1 MLAWVQAFLVSNMLAEAY.....PVDQEGSTPLMGAGTRPCA 263

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1393	100.0	263	4	Q96FE7	Q96FE7 homo sapien
2	1390	99.8	263	4	Q00318	Q00318 homo sapien
3	1379	99.0	263	4	Q8NCG9	Q8NCG9 homo sapien
4	167	12.0	562	6	Q8SQ23	Q8SQ23 sus scrofa
5	154.5	11.1	516	4	Q9BU99	Q9BU99 homo sapien
6	150	10.8	653	11	Q8VCS4	Q8VCS4 mus musculus
7	147.5	10.6	559	11	Q91VP2	Q91VP2 mus musculus
8	145.5	10.4	564	6	Q8MKR1	Q8MKR1 oryctolagus
9	142	10.2	810	4	Q15146	Q15146 homo sapien
10	139.5	10.0	385	5	Q25101	Q25101 herdamania m
11	137	9.8	716	11	P70521	P70521 rattus norv
12	136.5	9.8	420	13	Q90504	Q90504 epistretus
13	136	9.8	421	13	Q8AXX3	Q8AXX3 xenopus lae
14	135	9.7	90	4	Q8NG20	Q8NG20 homo sapien
15	135	9.7	812	11	Q9RW3	Q9RW3 rattus norv
16	134.5	9.7	395	4	Q9BZW1	Q9BZW1 homo sapien

17	134.5	9.7	704	13	Q90865	Q90865 gallus gall
18	134	9.6	716	11	Q91XG8	Q91XG8 mus musculus
19	131.5	9.4	313	13	Q9PU78	Q9PU78 crocodylus
20	131	9.4	154	4	Q96SE8	Q96SE8 homo sapien
21	131	9.4	608	11	Q9PTW7	Q9PTW7 struthio ca
22	130.5	9.4	461	11	Q8K1S7	Q8K1S7 mus musculus
23	130.5	9.4	717	13	P70006	P70006 xenopus lae
24	129	9.3	616	6	Q97507	Q97507 sus scrofa
25	128.5	9.2	806	6	Q18783	Q18783 macropus eu
26	127	9.1	157	6	Q9TV48	Q9TV48 bos taurus
27	126.5	9.1	433	6	Q8M1L0	Q8M1L0 oryctolagus
28	126.5	9.1	433	6	Q8MHY7	Q8MHY7 oryctolagus
29	126.5	9.1	728	11	Q8C9G5	Q8C9G5 mus musculus
30	126	9.0	716	13	Q91691	Q91691 xenopus lae
31	125.5	9.0	560	4	Q14520	Q14520 homo sapien
32	124	8.9	429	13	Q8AVB0	Q8AVB0 brachydanio
33	123	8.8	728	6	Q9BH09	Q9BH09 felis silve
34	121.5	8.7	399	4	Q96GL8	Q96GL8 homo sapien
35	121.5	8.7	420	4	Q9BTP9	Q9BTP9 homo sapien
36	121.5	8.7	424	4	Q8NCW1	Q8NCW1 homo sapien
37	121.5	8.7	462	4	Q8NCW0	Q8NCW0 homo sapien
38	121	8.7	381	4	Q8N2J4	Q8N2J4 homo sapien
39	120	8.6	231	11	Q8C6L2	Q8C6L2 mus musculus
40	119	8.5	334	6	Q46507	Q46507 papio hamad
41	119	8.5	710	13	Q91402	Q91402 xenopus. lae
42	118.5	8.5	615	4	Q81Z25	Q81Z25 homo sapien
43	117.5	8.4	812	11	Q91W05	Q91W05 mus musculus
44	114.5	8.2	113	4	Q9UIR7	Q9UIR7 homo sapien
45	114	8.2	113	4	Q9UIR5	Q9UIR5 homo sapien

#### ALIGNMENTS

RESULT 1  
Q96FE7 PRELIMINARY: PRT: 263 AA.

AC Q96FE7; 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; BC011049; AAH11049.1; -  
DR InterPro; IPRO00001; Kringle.  
DR Pfam; PF00051; Kringle; 1.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Hypothetical protein; Glycoprotein; Kringle.  
SQ  
SEQUENCE 263 AA; 28234 MW; 197C3EB888FA242 CRC64;

Query Match 100.0%; Score 1393; DB 4; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1.9e-120;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAWVQAFLVSNMLAEAYSGGCFWNGHLYREDQTSAPGLRCINWIDAGSGLASAP 60  
DB 1 MLAWVQAFLVSNMLAEAYSGGCFWNGHLYREDQTSAPGLRCINWIDAGSGLASAP 60  
QY 61 VSAGNHSYRNDEDPGRWCYVSGEAGPEKRPCEDLRCPTTSQALAFTEIOEAS 120



```
Db 61 VSGAGNHSYCRNDEDEPRGWCYVSGEAGVPEKRPEDIRCPETTSQALPAFTTEIOEAS 120
QY 121 EGGADDEVQVFAFAPNALPARSEAAAVQPVIGISQVRMNSKEKKDGLTGLVGLITMMVI 180
DB 121 EGGADDEVQVFAFAPNALPARSEAAAVQPVIGISQVRMNSKEKKDGLTGLVGLITMMVI 180
QY 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTEIIVDEKTVVHT 240
DB 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTEIIVDEKTVVHT 240
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263

RESULT 2
000318 PRELIMINARY; PRT; 263 AA.
ID 000318;
AC 000318;
DT 01-JUN-1997 (TrEMBLrel. 04, Created)
DT 01-JUN-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE WGSQ:DJ515N1.2 protein.
GN WGSQ:DJ515N1.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Du Z., Scheet P., Harper M.;
RT "The sequence of H. sapiens PAC clone RP3-515N1."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AC002073; AAB54054.1; -.
DR HSP; P00749; IKDU.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE; FALSE_NEG.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
KW Glycoprotein; Kringle.
SQ SEQUENCE 263 AA; 28248 MW; 197C3EE8E54A242 CRC64;

Query Match 99.8%; Score 1390; DB 4; Length 263;
Best Local Similarity 99.6%; Pred. No. 3.6e-120;
Matches 262; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 3
08NCJ9 PRELIMINARY; PRT; 263 AA.
ID 08NCJ9;
AC 08NCJ9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ90207.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isogai T., Ota T., Nishikawa T., Hayaashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Makamatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuko Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Nimomiyu K.;
RT "NEBO human cDNA sequencing project."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AK074688; BC11140.1; -.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
KW Hypothetical protein; Glycoprotein; Kringle.
SQ SEQUENCE 263 AA; 28104 MW; 73A9294D5142C63 CRC64;

Query Match 99.0%; Score 1379; DB 4; Length 263;
Best Local Similarity 99.2%; Pred. No. 3.8e-119;
Matches 261; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLAMVQAFVSNMLLAAYSGGCFWMDNGHLYREDQTSAPGRLCLNWLDAQGLASAP 60
DB 1 MLAMVQAFVSNMLLAAYSGGCFWMDNGHLYREDQTSAPGRLCLNWLDAQGLASAP 60
QY 61 VSGAGNHSYCRNDEDEPRGWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120
DB 61 VSGAGNHSYCRNDEDEPRGWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120
QY 121 EGGADDEVQVFAFAPNALPARSEAAAVQPVIGISQVRMNSKEKKDGLTGLVGLITMMVI 180
DB 121 EGGADDEVQVFAFAPNALPARSEAAAVQPVIGISQVRMNSKEKKDGLTGLVGLITMMVI 180
QY 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTEIIVDEKTVVHT 240
DB 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTITPLSAFTNPTEIIVDEKTVVHT 240
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263

RESULT 4
08SQ23 PRELIMINARY; PRT; 562 AA.
ID 08SQ23;
AC 08SQ23;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE T-Plasminogen activator.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
```

```

Query Match Similarity      12.0%: Score 167; DB 6; Length 562;
Best local similarity      36.6%: Pred No. 8,5e-07;
Matches      41; Conservative      12; Mismatches      35; Indels      26; Gaps      6

QY      25 CFWDNGHLYREDQTSFAPRLCLNMLDAOSGLASAPVS-----GAGNHSYCRNDE 75
Db      127 CYEDQGITRYGTWSTTSGAECVNM--NTSGIASPMYNGRRPDVAKYLGJGNNHYCRNPDK 184
QY      76 DPGRWCVV-SGEAGVPEKRPEDRCPEFTTSQALPATTEIQEASBEGPAD 126
Db      185 DSK-FWCYIFRAEKYSPD-----FC-----STPACTKEKECYTKGID 222

RESULT 5
Q9BU99          PRELIMINARY;      PRT;      516 AA.
AC      Q9BU99;
DT      01-JUN-2001 (Tremblrel. 17, Created)
DT      01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT      01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE      Similar to plasminogen activator, tissue.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_Taxid=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Skin;
RA      Strausberg R.;
RL      Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC      -!- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR      EMBL: BC002795; AA02795.1; -.
DR      HSSP: P00750.1ASH.
DR      InterPro: IPR001314, Chymotrypsin.
DR      InterPro: IPR006209, EGF-like.
DR      InterPro: IPR006210, IEGF.
DR      InterPro: IPR000001, Kringle.

```

Query Match	11.1%	Score 154.5	DB 4	Length 516
Best Local Similarity	39.6%	Pred. 1.1e-05		
Matches	36	Conservative	6	Mismatches 36; Indels 13; Gaps 4
QY	25	CEFMNGHYRDRQSSPAFLRCLTMMDLAQSLAAPV-----GAGHSHYRNRDE	75	
DB	81	CYEDQGISRYRTWTSTASGAECTNM--NSSLAQKPSGRPRDAIRLGSHNNYCRNDR	138	
QY	76	DPRGPWCYVSGEAGVPEKRPCCEDLRCPETTS	106	
DB	139	DSK--PWCYVF--KAGKYSGEEFCSTPACSGNS	167	

```

RESULT 6
Q8VCS4          PRELIMINARY;          PR7;          653 AA.
ID Q8VCS4;
AC Q8VCS4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical 70.6 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; BC019376; AAH19376.1; -.
DR HSSP; P00761; 1AN1.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000083; Fibriectn.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00013; ENTPB11.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000995; FN_Type_II; 1.
DR ProDom; PD000395; Kringle; 1.

```



Query Match 10.4%; Score 145.5; DB 6; Length 564;  
Best Local Similarity 39.5%; Pred. No. 8.2e-05;  
Matches 34; Conservative 7; Mismatches 32; Indels 13; Gaps 4;

QY 25 CPWNGHLYREDQTSFAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRG 79  
DB 128 CYEDRGIGYRGFTWSTTESGAQCVMN--NSSWLALPKYSGRKNALRLGLGNHYCRNDR 185  
QY 76 DPGPMCVYSGVAGEPEKRPCEDLRC 101  
DB 186 DTK-PWCYVF-RAGIYSPFCSTPAC 209

RESULT 9  
ID Q25101 PRELIMINARY; PRT; 810 AA.  
AC Q25101;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Plasminogen precursor.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Browne M.J., Chapman C.G., Dodd I., Carey J.E., Lawrence G.M.P.,  
RA Mitchell D., Robinson J.H.;  
RT "Expression of recombinant human plasminogen and aglycoplasminogen in  
Hela cells.";  
RL Fibrinolysis 0:0-0(1991).  
CC -I- SIMILARITY: CONTAINS 5 KRINGLE DOMAINS.  
DR EMBL; M74220; AAA56451.1; -.  
DR HSSP; P00747; 2PK4.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR003014; PAN.  
DR InterPro; IPR003609; Pan\_app.  
DR InterPro; IPR003966; Prothrombin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00051; kringle; 5.  
DR Pfam; PF00024; PAN; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPsin.  
DR PRINTS; PR00018; KRINGLE.  
DR PRINTS; PR01505; PROTHROMBIN.  
DR ProDom; PD000395; Kringle; 5.  
DR SMART; SM00130; KR; 5.  
DR SMART; SM00473; PAN\_AP; 1.  
DR SMART; SM00020; TRYp\_Spc; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 5.  
DR PROSITE; PS50070; KRINGLE\_2; 5.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
KW Glycoprotein; Hydrolyase; Kringle; Protease; Serine protease; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 810 PLASMINOGEN.  
SQ SEQUENCE 810 AA; 90555 MW; B05CTD4B0D020B3C CRC64;

Query Match 10.2%; Score 142; DB 4; Length 810;  
Best Local Similarity 37.2%; Pred. No. 0.00028;  
Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4;

QY 25 CPWNGHLYREDQTSFAPGLRCLNWLDAQSGLASAPVSGAGNHSYCRNPDEDPRG 79  
DB 103 CKTGNGKNGYGTWSTTESGAQCVMN--NSSWLALPKYSGRKNALRLGLGNHYCRNDR 185  
QY 80 PWCYSGVAGEPEKRPCEDLRCPE 103

DB 162 PWCYTTD----PEKRYDCDILBCEE 183

RESULT 10  
ID Q25101 PRELIMINARY; PRT; 385 AA.  
AC Q25101;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Serine proteinase.  
GN HMERPI.  
OS Herdmania momus.  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Stolidobranchia; Pyuridae; Herdmania.  
OX NCBI\_TaxID=7733;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=curvata;  
RA Arnold J.M., Kennet C., Lavin M.P.;  
RT "Transient expression of a novel serine protease in the ectoderm of  
the ascidian Herdmania momus during development.";  
RL Dev. Genes Evol. 206:455-463(1997).  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -I- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; U63517; AAB6650.1; -.  
DR HSSP; P00763; IDPO.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00051; kringle; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPsin.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00020; TRYp\_Spc; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS50070; KRINGLE\_2; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
KW Hydrolyase; Kringle; Protease; Serine protease.  
SQ SEQUENCE 385 AA; 42935 MW; BFB1D05D532B6A0 CRC64;

Query Match 10.0%; Score 139.5; DB 5; Length 385;  
Best Local Similarity 20.2%; Pred. No. 0.00018;  
Matches 70; Conservative 47; Mismatches 127; Indels 103; Gaps 14;

QY 2 LIAM--VQAFVLSNMLABAYSGGCF-WDNGHLYREDQTSFAPGLRCLNWLDAQSG 54  
DB 7 LVIWIIAGFVESN-----SBCFDLENESYOGAISRLLGGETCGSW-DLQTPKH 56  
QY 55 --GLSAPVSGAGNHSYCRNPDEDPRGPMCVYSGE-----AGPEKRPCEDLR 100  
DB 57 KTTSSNYPVSGAAGNHYCRNPDPDRGPMCVYTTNEMRMWDYCDIPICSNPPVTLPPSIE 116  
QY 101 CPEET--SQALPAFTTEIOEASSEP-----GADEVQ----- 129  
DB 117 CGKTTPEPLSDATKGYGLKSKAKTNLHVTGTTVHGSIPMVSRLKRELHFGGSI 176  
QY 130 ----VFAPANALPARSEAAVOPVIGISQRYRANSEKKKDLGLTVL----- 173  
DB 177 LNRNWLITLAHCHIRKQCPKYLALIGDYDRIOYDPESEK-----VGFRLFNHEKYNPAT 232  
QY 174 ---GIMMYTIIAIGAILGYSYKRGKDLKEGHDQVC-----EREMORITL 218  
DB 233 FENDITLKKDQDTSIATITFGQSVFPPANVPAKSKIIYSGMGDTKGTQDYKLVNVTL 292  
QY 219 PLSAFTNPCEIVDEKTV-----VWTSQTPVVDQEGSTPLMGAGTP 261  
DB 293 PVMSE--KLCKRLYSKVVGAAVFKTSLCAAYKKGKGDSCGDSGGP 337

RESULT 11  
P70521  
ID P70521 PRELIMINARY; PRT; 716 AA.  
AC P70521;  
DT 01-FEB-1997 (TRMBLrel. 02, Created)  
DT 01-FEB-1997 (TRMBLrel. 02, Last sequence update)  
DT 01-MAR-2003 (TRMBLrel. 23, Last annotation update)  
DE Macrophage stimulating protein precursor.  
CN MSP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=97011126; PubMed=8858136;  
RA Ohashiro K., Iwama A., Matsuno K., Ezaki T., Sakamoto O., Hamaguchi I.,  
RA Takasu N., Suda T.;  
RT "Molecular cloning of Rat Macrophage-stimulating protein and its  
RT involvement in the Male Reproductive System";  
RT Biochem Biophys. Res. Commun. 227:273-280(1996).  
CC -1- SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.  
DR EMBL: X95096; CAA64473.1; -.  
DR HSSE; P00747; 1KRN.  
DR MEROPS; S01.975; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR003014; PAN.  
DR InterPro; IPR003609; Pan\_app.  
DR InterPro; IPR003966; Prothrombin.  
DR Pfam; PF00051; Kringle; 4.  
DR Pfam; PF00024; PAN; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR PRINTS; PR01505; PROTHROMBIN.  
DR PRODOM; PD000395; Kringle; 4.  
DR SMART; SM00130; KR; 4.  
DR SMART; SM00473; PAN AP; 1.  
DR SMART; SM00020; Tryp\_Spec; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 4.  
DR PROSITE; PS50070; KRINGLE\_2; 4.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
KW Glycoprotein; Hydrolase; Kringle; Protease; Serine protease; Signal.  
FT SIGNAL 1  
FT SIGNAL 31  
SQ SEQUENCE 716 AA; 80733 MW; 06B7DF3BFE6D21F CRC64;

Query Match 9.8%; Score 137; DB 11; Length 716;  
Best Local Similarity 27.3%; Pred. No. 0.0069;  
Matches 48; Conservative 12; Mismatches 56; Indels 60; Gaps 8;

QY 2 LLAWQAFVLSNMLAAEAVGSG-----GCFMNGHLVREDQTSAPGLRCLNW 49  
DB 80 LHPMTQ-----HSLAQLHHSGLCDLFQKDYVTCIMDNQASIRGVIARADGUPCOAM 134  
QY 50 ---LDAQSLASAPVSGAGNHSYCRNPDEDPRGPWCYVS----- 85  
DB 135 SRPFNDHKYPTTPKNGL-EEVFCGNPDGDPFGPCYTTNSVRFQSGIKSCREAVCW 193  
QY 86 -----GEAGVPER-RPCE--DIACPEF-----TSQALPAFTTIQASRGP 123  
DB 194 CNGEDYRGSEVDVTESGREGQKWDLQHPHSHPEKFPDKALKONYCRNPDSERP 249

RESULT 12  
Q90504  
ID Q90504 PRELIMINARY; PRT; 420 AA.  
AC Q90504;  
DT 01-NOV-1996 (TRMBLrel. 01, Created)  
DT 01-NOV-1996 (TRMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TRMBLrel. 23, Last annotation update)  
DE Thrombin.  
OS Eptaretus stouiti (Pacific hagfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;  
OC Myxiniidae; Eptaretinae; Eptaretus.  
OX NCBI\_TaxID=7765;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=92212913; PubMed=1557383;  
RA Banfield D.K., Macgillivray R.T.;  
RT "Partial characterization of vertebrate prothrombin cDNAs:  
RT amplification and sequence analysis of the B chain of thrombin from  
RT nine different species";  
RT Proc. Natl. Acad. Sci. U.S.A. 89:2779-2783(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=944223694; PubMed=7513365;  
RA Banfield D.K., Irwin D.M., Walz D.A., Macgillivray R.T.;  
RT "Evolution of prothrombin: isolation and characterization of the cDNAs  
RT encoding chicken and hagfish prothrombin";  
RT J. Mol. Evol. 38:177-187(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Banfield D.K.;  
RT Submitted (DEC-1991) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL: M61393; AAA21620.1; -.  
DR HSSE; P00734; 1UVS.  
DR MEROPS; S01.217; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR003966; Prothrombin.  
DR InterPro; IPR001254; Ser. protease\_Try.  
DR Pfam; PF00051; Kringle; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR PRINTS; PR01505; PROTHROMBIN.  
DR PRODOM; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00020; Tryp\_Spec; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS50070; KRINGLE\_2; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Kringle; Protease; Serine protease.  
FT SIGNAL 420 AA; 47888 MW; 64522AA21A57B67A CRC64;

Query Match 9.8%; Score 136.5; DB 13; Length 420;  
Best Local Similarity 26.5%; Pred. No. 0.0038;  
Matches 39; Conservative 16; Mismatches 77; Indels 15; Gaps 4;

QY 25 CFMNGHLVREDQTSAPGLRCLNWLDQAQSLASAPVSGAG-NHSYCRNPDEDPRGPWCY 83  
DB 17 CYARRERDRYKGLNLTWTGKPCLPWRGSYNLPQQTFTAGLTNSYCRNPDDSDSGWCY 76  
QY 84 VSSEAGVP-----EKRCEDLRCPETTSQALPAFTTEIOBASGPADEVOVAPANALPA 139  
DB 77 TKVEGTVDVYQQLNVC-----SGDIFVEGTDEVQLSGSBSGAETTLFPNKTGON 129  
QY 140 RSEFAAVQPVIGISQVRNNSKEKKDL 166  
DB 130 GEEECGRKRPWFLOKQ--NDRSEDEL 153

RESULT 13  
Q8AXX3  
ID Q8AXX3 PRELIMINARY; PRT; 421 AA.  
AC Q8AXX3;

DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Kremen2.  
GN KRM2.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8335;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22308873; PubMed=12421700;  
RA Davidson G., Mao B., Del Barco Barrantes I., Niehs C.;  
RT "Kremen proteins interact with Dickkopf1 to regulate anteroposterior  
RT CNS patterning."  
RL Development 129:5587-5596(2002).  
DR EMBL; AY150813; AAN64661.1;  
SQ SEQUENCE 421 AA; 46274 MW; 74D04DB1682CD15F CRC64;

Query Match 9.8%; Score 136; DB 13; Length 421;  
Best Local Similarity 43.8%; Pred. No. 0.0042;  
Matches 32; Conservative 5; Mismatches 28; Indels 8; Gaps 4;

OY 25 CFWDGCHLYREDQTSAP-GLRCLMWLDASGLASAPV-----GAGNHSYCRNPDEDP 78  
DB 29 CFTVAGRDYRGTVSAGBPGTPLYNQTTQHLVNAQSDPDGELGLNHNVCRNPDADVQ 88  
OY 79 GPCVYS-GEAGV 90  
DB 89 -PWCYSENEBGI 100

## RESULT 14

08NG20 PRELIMINARY; PRT; 90 AA.  
AC 08NG20;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Plasmidogen/activator kringle.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominda; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dou D.;  
RT "Production of kringle fragment."  
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
DR EMBL; AF282882; AAM52248.1;  
DR InterPro; IPR000001; Kringle.  
DR Pfam; PF00051; kringle\_1.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
KW Glycoprotein; Kringle.  
SQ SEQUENCE 90 AA; 9804 MW; A33887F9DF4C7B1 CRC64;

Query Match 9.7%; Score 135; DB 4; Length 90;  
Best Local Similarity 37.8%; Pred. No. 6.8e-05;  
Matches 28; Conservative 8; Mismatches 26; Indels 12; Gaps 3;

OY 21 GSGGCFWDNGHLYREDQTSAPGLRCLNW-----LDAQGLASAPVSGAGNHSYCR 71  
DB 4 GNSDCYFGNGSAYRGTHSLTSGASCLPWNMILICKVYTAQN--PSAALGLGKHNYCR 61  
OY 72 NPDEDRGPWCYVS 85  
DB 62 NPDGAK-PWCYTT 74

## RESULT 15

09ROW3 PRELIMINARY; PRT; 812 AA.  
AC 09ROW3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Plasmidogen protein precursor (EC 3.4.21.7).  
GN PLASMINOGEN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Bangert K., Johnsen A.H., Thorsen S.;  
RT "Rat plasminogen cDNA and gene structure."  
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=91250378; PubMed=1645711;  
RC TISSUE=Liver;  
RA Karlas J.J., Makker S.P.;  
RT "Identification of the rat Heymann nephritis autoantigen (GP330) as a  
RT receptor site for plasminogen."  
RL J. Biol. Chem. 266:10825-10829(1991).  
CC -1- SIMILARITY: CONTAINS 5 KRINGLE DOMAINS.  
DR EMBL; AJ242649; CAB46014.1;  
DR HSRP; P00747; 1PKK.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR000001; kringle.  
DR InterPro; IPR003014; PAN.  
DR InterPro; IPR003609; Pan app.  
DR InterPro; IPR003966; Prothrombin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR InterPro; IPR001400; Somatostatin.  
DR Pfam; PF00051; kringle; 5.  
DR Pfam; PF00024; PAN; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR PRINTS; PR01505; PROTHROMBIN.  
DR ProDom; PD000395; Kringle; 5.  
DR SMART; SM00130; KR; 4.  
DR SMART; SM00473; PAN AP; 1.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 5.  
DR PROSITE; PS00070; KRINGLE\_2; 5.  
DR PROSITE; PS00338; SOMATOSTATIN\_2; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Glycoprotein; Hydrolase; Kringle; Protease; Serine protease; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 812 PLASMINOGEN.  
SQ SEQUENCE 812 AA; 90535 MW; 8C703C51410EBC9E CRC64;

Query Match 9.7%; Score 135; DB 11; Length 812;  
Best Local Similarity 30.9%; Pred. No. 0.0012;  
Matches 34; Conservative 14; Mismatches 48; Indels 14; Gaps 5;

OY 25 CFWDGCHLYREDQTSAPGLRCLMWLD-----QGLASAPVSGAGNHSYCRNPDEDP 79  
DB 376 CTAQNGKSTKRGSSSTNTCKKQSVSMTPHSHSKTPANFPAGL-ENMYCNPDPDQNG 434  
OY 80 PWCYSGEAGVPEKR--PCEDLRCPETTSQALPAFTTEIQASBEGGADE 127  
DB 435 PWCFTTD-----PSVWEYCNLRKCSGTGGV--AESAIIVQVPSABGTSE 478

Search completed: November 25, 2003, 13:25:33  
Job time : 42 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 12:21:02 ; Search time 41 Seconds

(without alignments)  
1018.173 Million cell updates/sec

Title: US-10-057-951-2

Sequence: 1 MLLAWVQAFIVSNMLLEAY.....PVDPQEGSTPLMGQAGTPGA 263

Scoring table: BLOSUM62

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
```

Database :

A\_Geneseq.10Jun03.\*

1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*

2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*

3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*

4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*

5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*

6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*

7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*

8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*

9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*

10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*

11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*

12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*

13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*

14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*

15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*

16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*

17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*

18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*

19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*

20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*

21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*

22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*

23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1393	100.0	263	20	AAV05219	Kirngliel protein s
2	1393	100.0	263	20	AAW87769	Human tissue plas
3	1393	100.0	263	22	AAE00300	Human tissue-plas
4	1393	100.0	263	23	ABR40414	Human secreted pro
5	1390	99.4	263	23	AAU86149	Human PRO264 poly
6	1384	99.4	263	21	AAU84337	Human ORF301Y
7	1379	99.0	263	22	AAW93748	Human polypeptide,
8	1167.5	83.8	286	20	AAV05220	Kirngliel protein s
9	752	54.0	146	23	ABR40487	Human secreted pro

10	752	54.0	146	23	ABR40561	Human secreted pro
11	322	23.1	66	22	ABG52752	Human liver peptid
12	322	23.1	66	22	ABR37995	Peptide #5411 enco
13	322	23.1	66	22	ABR33159	Protein #5158 enco
14	322	23.1	66	22	AAM58537	Human Brain expres
15	322	23.1	66	22	AAM71037	Human bone marrow
16	322	23.1	66	22	AAM18900	Peptide #5234 enco
17	322	23.1	66	23	AAM11314	Peptide #5351 enco
18	322	23.1	66	23	ABG40828	Human peptide enco
19	306	22.0	56	20	AAV12615	Human 5' EST secret
20	305	21.9	55	20	AAV12337	Human 5' EST secret
21	196	14.1	39	19	AAW72641	Nervous glia cell
22	192	13.8	39	19	AAW26240	Nervous glia cell
23	160.5	11.5	527	13	AAAR02127	t-PA analogue expr
24	160.5	11.5	527	13	AAAR02230	t-PA analogue expr
25	160.5	11.5	527	13	AAAR02128	t-PA analogue expr
26	160.5	11.5	527	13	AAAR02129	t-PA analogue expr
27	160.5	11.5	593	8	AAAP70059	Sequence of hybrid
28	160.5	11.5	650	8	AAAP70060	Sequence of hybrid
29	160	11.5	351	20	AAAP25407	Human tissue factor
30	159.5	11.5	472	10	AAAP94416	Sequence of coding
31	159.5	11.5	527	13	AAAR02223	t-PA analogue expr
32	159.5	11.5	527	13	AAAR02222	t-PA analogue expr
33	159.5	11.5	527	13	AAAR02021	t-PA analogue expr
34	157.5	11.3	527	19	AAAM54154	t-PA mutant (N1422
35	157.5	11.3	527	19	AAAM54157	t-PA mutant (N1422
36	157.5	11.3	562	10	AAAP94379	Amino acid sequenc
37	157.5	11.3	562	10	AAAP94380	Amino acid sequenc
38	156.5	11.2	439	16	AAAR68851	Delta 2-89 tissue
39	156.5	11.2	483	16	AAAR70889	Human tissue PA vs
40	156.5	11.2	483	16	AAAR70877	Human tissue PA vs
41	156.5	11.2	483	16	AAAR70878	Human tissue PA vs
42	156.5	11.2	483	16	AAAR70879	Human tissue PA vs
43	156.5	11.2	483	16	AAAR70880	Human tissue PA vs
44	156.5	11.2	483	16	AAAR70881	Human tissue PA vs
45	156.5	11.2	483	16	AAAR70882	Human tissue PA vs

## ALIGNMENTS

RESULT 1	
AA050219	
ID	AA050219 standard; Protein; 263 AA.
AC	AA050219;
XX	
DT	17-JUN-1999 (first entry)
XX	
DE	Kringel protein sequence.
XX	
KW	Kringel; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS
KW	CNS inflammation; cerebellar degeneration; Alzheimer's disease; ashma;
KW	Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;
KW	neurological abnormality; ischaemia reperfusion injury; ischaemic injury
KW	cardiovascular disease; kidney disease; liver disease; aplastic anaemia;
KW	myocardial infarction; hypotension; hypertension; allergy; infection;
KW	myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;
XX	male pattern baldness.
XX	
OS	Homo sapiens.
XX	
PN	W09911788-A1.
XX	
PD	11-MAR-1999.
XX	
PF	02-SEP-1998; 98WO-US18270.
XX	
PR	01-SEP-1998; 98US-014889.
XX	
PR	02-SEP-1997; 97US-0056032.
XX	
PA	(SMIK ) SMITHKLINE BEECHAM CORP.
XX	

PI Albore EF, Kikly KK;  
 XX WPI: 1999-214707/18.  
 DR N-PSDB; AAX28354.  
 XX  
 PT New Kringle1 polypeptides and polynucleotides  
 XX  
 PS Claim 1; Page 31-32; 42pp; English.  
 XX  
 CC This sequence is a Kringle1 polypeptide of the invention.  
 CC The Kringle1 polypeptides (I) are used to screen for agonists and  
 CC antagonists. Agonists are used to treat subjects in need of enhanced  
 CC activity or expression of (I). Antagonists are used to treat subjects  
 CC having need to inhibit the activity or expression of (I). The methods can  
 CC be used to treat conditions such as cancer, inflammation, autoimmunity,  
 CC allergy, asthma, rheumatoid arthritis, CNS inflammation, cerebellar  
 CC degeneration, Alzheimer's disease, Parkinson's disease, multiple  
 CC sclerosis, amyotrophic lateral sclerosis, head injury damage and other  
 CC neurological abnormalities, ischaemia reperfusion injury, cardiovascular  
 CC disease, kidney disease, liver disease, ischaemic injury, myocardial  
 CC infarction, hypotension, hypertension, AIDS, myelodysplastic syndromes  
 CC and other haematologic abnormalities, aplastic anaemia, male pattern  
 CC baldness, and bacterial, fungal, protozoan and viral infections. The  
 CC Kringle1 polypeptides may also be used to generate antibodies.  
 CC Determining the presence or absence of mutations in, and analysing for  
 CC the presence or absence of expression of, Kringle1 polynucleotides can be  
 CC used to diagnose a disease or susceptibility to a disease related to  
 CC expression or activity of Kringle1 proteins. The polynucleotides may also  
 CC be used for chromosome identification, and mapping.  
 CC  
 XX  
 SQ Sequence 263 AA;  
 Query Match 100.0%; Score 1393; DB 20; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 8,3e-120;  
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLTAMVQAFIVSNMLLEAVYSGGCFMDNGHLYREDQTPAPGRCINWLDAGSLASAP 60  
 DB 1 MLTAMVQAFIVSNMLLEAVYSGGCFMDNGHLYREDQTPAPGRCINWLDAGSLASAP 60  
 QY 61 VSGAGNHSYCNRPEDPRGWCYVSGENGVEKPCPELRCPETTSQALPFTTEIQEAS 120  
 DB 61 VSGAGNHSYCNRPEDPRGWCYVSGENGVEKPCPELRCPETTSQALPFTTEIQEAS 120  
 QY 121 EGPAGDEVQVAPAPALPARSEAAAVCPVIGISGRVRSNKKKDLGLGYVLGITMMVI 180  
 DB 121 EGPAGDEVQVAPAPALPARSEAAAVCPVIGISGRVRSNKKKDLGLGYVLGITMMVI 180  
 QY 181 IIAICAGIILGYSYKRGDLXKOHDKVCEREMORITPLSAFTNPTCEIIDEKTVVHT 240  
 DB 181 IIAICAGIILGYSYKRGDLXKOHDKVCEREMORITPLSAFTNPTCEIIDEKTVVHT 240  
 QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263  
 DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263  
 RESULT 2  
 AAW87769  
 ID AAW87769 standard; Protein; 263 AA.  
 XX  
 AC AAW87769;  
 XX  
 DT 29-MAR-1999 (first entry)  
 XX  
 DE Human tissue plasminogen activator-like protease t-PALP.  
 XX  
 KW Tissue plasminogen activator-like protease; t-PALP; human;  
 KW circulatory system-related disorder; blood clotting; stroke;  
 KW thrombosis; peripheral arterial occlusion; pulmonary embolism;  
 KW myocardiothrombosis; diagnosis; therapy.  
 XX  
 OS Homo sapiens.

XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /label= Sig\_peptide  
 FT Protein 22..263  
 FT /label= Mat\_protein  
 FT Domain 25..84  
 FT /note= "kringle domain"  
 FT Domain 85..263  
 FT /note= "protease domain"  
 FT Peptide 22..31  
 FT /note= "epitope-bearing region"  
 FT Peptide 35..44  
 FT /note= "epitope-bearing region"  
 FT Peptide 71..81  
 FT /note= "epitope-bearing region"  
 FT Peptide 91..107  
 FT /note= "epitope-bearing region"  
 FT Peptide 119..128  
 FT /note= "epitope-bearing region"  
 FT Peptide 138..147  
 FT /note= "epitope-bearing region"  
 FT Peptide 155..167  
 FT /note= "epitope-bearing region"  
 FT Peptide 193..203  
 FT /note= "epitope-bearing region"  
 FT Peptide 206..215  
 FT /note= "epitope-bearing region"  
 FT Peptide 227..237  
 FT /note= "epitope-bearing region"  
 FT Peptide 243..252  
 FT /note= "epitope-bearing region"  
 XX  
 PN WO9854199-A1.  
 XX  
 PD 03-DEC-1998.  
 XX  
 XX 27-MAY-1998; 98WO-US10728.  
 XX  
 PR 28-MAY-1997; 97US-0048000.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ebner R, Moore PA, Ruben SM,  
 XX  
 DR WPI: 1999-070207/06.  
 DR N-PSDB; AAV99636.  
 XX  
 PT New tissue plasminogen activator-like protease - useful in the  
 PT diagnosis and treatment of circulatory system-related disorders  
 PS Claim 1; Page 56-57; 76pp; English.  
 XX  
 CC This is the amino acid sequence of tissue plasminogen activator-like  
 CC protease (t-PALP), a novel member of the serine protease family  
 CC that shares sequence homology to human tissue plasminogen activator  
 CC (see AAW87770). The t-PALP sequence was deduced from a cDNA clone  
 CC (see AAV99636) derived from activated monocytes. The 2.5 kb t-PALP  
 CC message has also been detected in heart, brain, lung, placenta,  
 CC liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate,  
 CC testis, ovary, small intestine, colon and peripheral blood  
 CC leukocytes. Isolated nucleic acids encoding amino acids -21 to  
 CC 242, -20 to 242, 1-242, 4-63 (Kringle domain) and 64-242 (protease  
 CC domain) of t-PALP, or encoding epitope-bearing portions of t-PALP,  
 CC are also claimed, as are recombinant vectors, host cells, and  
 CC methods for producing t-PALP polypeptides. t-PALP may be used to  
 CC detect and treat disorders related to the circulatory system, and  
 CC to identify agonists and antagonists of t-PALP activity. The  
 CC homology between t-PALP and tPA indicates that t-PALP may be  
 CC involved in the regulation of normal and abnormal clotting  
 CC in e.g. stroke, deep-vein thrombosis, peripheral arterial  
 CC occlusion, pulmonary embolism and myocardiothrombosis.

SQ Sequence 263 AA;  
 Query Match 100.0%; Score 1393; DB 20; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 8,3e-120;  
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLIAVQALVLSNMLAEAYSGGCFWNDGHIYREDQTSAPAGLACIMNLDAGSLASAP 60  
 1 MLIAVQALVLSNMLAEAYSGGCFWMDGHIYREDQTSAPAGLACIMNLDAGSLASAP 60  
 DB 1 VSAGNHSYCRNPDEDPGRCVSGEAVPEKRCPCEDRCETTSOALPAPFTBIQENS 120  
 61 VSAGNHSYCRNPDEDPGRCVSGEAVPEKRCPCEDRCETTSOALPAPFTBIQENS 120  
 QY 121 EBPGEDEVQVFPANALPARSEAAVQPIGISQVRNNSKEKDLGTLGYVIGITMVI 180  
 121 EBPGEDEVQVFPANALPARSEAAVQPIGISQVRNNSKEKDLGTLGYVIGITMVI 180  
 DB 121 EBPGEDEVQVFPANALPARSEAAVQPIGISQVRNNSKEKDLGTLGYVIGITMVI 180  
 181 IIAIAGAILGYSYRKQDLKEQHDQKVCEREMORTLPLSFNPTCEIVDEKIVVHT 240  
 181 IIAIAGAILGYSYRKQDLKEQHDQKVCEREMORTLPLSFNPTCEIVDEKIVVHT 240  
 QY 241 SOTVPDQGSTPLMGQAGTPGA 263  
 241 SOTVPDQGSTPLMGQAGTPGA 263  
 DB 241 SOTVPDQGSTPLMGQAGTPGA 263

RESULT 3  
 AAE00300  
 ID AAE00300 standard; Protein; 263 AA.  
 AC AAE00300;  
 XX  
 DT 13-JUN-2001 (first entry)  
 DE Human tissue-plasminogen activator-like protease (t-PALP).  
 XX  
 XX Human; tissue-plasminogen activator-like protease; t-PALP;  
 KW therapy; vascular disease; stroke; deep vein thrombosis; keloid; asthma;  
 KW arterial occlusion; blood coagulation disorder; cerebroprotective;  
 KW autoimmune system disorder; human immunodeficiency syndrome; cystostatic;  
 KW rheumatoid arthritis; graft-versus-host disease; thyroiditis; candida;  
 KW insulin dependent diabetes; inflammatory eye disease; allergy; antiviral;  
 KW cardiovascular disease; heart disease; arhythmia; myocardial ischemia;  
 KW hyperproliferative disorder; hypertrophic scar; neurological disease;  
 KW Creutzfeldt-Jakob syndrome; neurodegenerative disorder; antiinflammatory;  
 KW Alzheimer's disease; Parkinson's disease; immunosuppressive; antifungal;  
 KW infectious disease; drug screening; gene therapy; neuroprotective;  
 KW cancer; ophthalmological; antibacterial; vulnerary.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Binding-site 1..165  
 FT /note= "Binds to FLAG polypeptide to form  
 FT t-PALP-FLAG fusion protein"  
 FT Peptide 1..21  
 FT /label= Signal\_peptide  
 FT Domain 4..63  
 FT /label= Kringle\_domain  
 FT Region 12..21  
 FT /note= "Conserved region"  
 FT Protein 22..263  
 FT /note= "human mature tissue-plasminogen activator-like  
 FT protease (t-PALP); Binds to FLAG polypeptide to form  
 FT t-PALP-FLAG fusion protein"  
 FT Region 22..38  
 FT /note= "Conserved region"  
 FT Region 22..31  
 FT /note= "Epitope-bearing portion"  
 FT Region 35..44  
 FT /note= "Epitope-bearing portion"  
 FT Region 39..49

FT Region /note= "Conserved region"  
 FT 50..62 /note= "Conserved region"  
 FT Region 63..84 /note= "Conserved region"  
 FT Domain 64..242 /label= Protease\_domain  
 FT Region 71..81 /note= "Epitope-bearing portion"  
 FT Region 85..97 /note= "Conserved region"  
 FT Region 91..107 /note= "Epitope-bearing portion"  
 FT Region 100..118 /note= "Conserved region"  
 FT Region 119..128 /note= "Epitope-bearing portion"  
 FT Region 119..127 /note= "Conserved region"  
 FT Region 128..143 /note= "Conserved region"  
 FT Region 138..147 /note= "Epitope-bearing portion"  
 FT Region 146..163 /note= "Conserved region"  
 FT Region 155..167 /note= "Epitope-bearing portion"  
 FT Region 164..180 /note= "Conserved region"  
 FT Region 186..200 /note= "Conserved region"  
 FT Region 193..203 /note= "Epitope-bearing portion"  
 FT Region 201..220 /note= "Conserved region"  
 FT Region 206..215 /note= "Epitope-bearing portion"  
 FT Region 221..236 /note= "Conserved region"  
 FT Region 227..237 /note= "Epitope-bearing portion"  
 FT Region 237..248 /note= "Conserved region"  
 FT Region 243..252 /note= "Epitope-bearing portion"  
 FT Region 249..263 /note= "Conserved region"  
 FT  
 XX  
 PN MO200125252-A1.  
 XX  
 PD 12-APR-2001.  
 XX  
 PF 03-OCT-2000; 2000WO-US27239.  
 XX  
 PR 04-OCT-1999; 99US-0411977.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Moore PA, Ruben SM, Ebner R;  
 DR WPI, 2001-235402/24.  
 DR N-PSDB; AAD03460.  
 XX  
 XX New (gene encoding and antibody immunospecific for a)  
 PT tissue-plasminogen activator-like protease, useful for the diagnosis  
 PT and treatment of (cardio)vascular diseases, hyperproliferative  
 PT disorders, immune system disorders and cancers -  
 PS Claim 17; Fig 1; 323pp; English.  
 XX The present amino acid sequence is HMSIB42 clone human  
 CC tissue-plasminogen activator-like protease (t-PALP). The t-PALP  
 CC sequence and their (ant)agonists are useful for the diagnosis and

CC treatment of vascular diseases e.g. stroke, deep vein thrombosis and  
 CC arterial occlusion, blood coagulation disorders, (auto)immune system  
 CC disorders e.g. human immunodeficiency syndrome, rheumatoid arthritis,  
 CC graft-versus-host disease, thyroiditis, insulin dependent diabetes and  
 CC inflammatory eye disease, allergic reactions e.g. asthma, cardiovascular  
 CC diseases e.g. heart disease, arrhythmia and myocardial ischaemia,  
 CC hyperproliferative disorders, cancers, hypertrophic scars and keloids,  
 CC neurological diseases e.g. Creutzfeldt-Jakob syndrome, neurodegenerative  
 CC disease e.g. Alzheimer's disease and Parkinson's disease and infectious  
 CC are also useful for drug screening. The t-PALP nucleotides are useful as  
 CC chromosome markers and are involved in gene therapy.

XX Sequence 263 AA;

Query Match 100.0%; Score 1393; DB 22; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-120;  
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTAWQAFVSNMMLAEAYSGGCFWMDNGLYREDQTSAPGRLCLNMLDAOSGLASAP 60  
 Db 1 MLTAWQAFVSNMMLAEAYSGGCFWMDNGLYREDQTSAPGRLCLNMLDAOSGLASAP 60  
 QY 61 VSGAGNHSYCRNPDEDPGRWCYVSGAGVPEKRPCEDLPCPTTSQALPAFTTEIOEAS 120  
 Db 61 VSGAGNHSYCRNPDEDPGRWCYVSGAGVPEKRPCEDLPCPTTSQALPAFTTEIOEAS 120  
 QY 121 EGGADDEVQVAPAPNALPARSEAAAVQVIGISQRYMNSKEKKDLGTGLGVIGITMVI 180  
 Db 121 EGGADDEVQVAPAPNALPARSEAAAVQVIGISQRYMNSKEKKDLGTGLGVIGITMVI 180  
 QY 181 IIAIGAGIILIGYSYKRGKDKLKEQHDQVCEREMQRTLP.SAFTNPCEIYDEKTVVHT 240  
 Db 181 IIAIGAGIILIGYSYKRGKDKLKEQHDQVCEREMQRTLP.SAFTNPCEIYDEKTVVHT 240  
 QY 241 SQTPVDPQEGSTPLMGAGTPGA 263  
 Db 241 SQTPVDPQEGSTPLMGAGTPGA 263

#### RESULT 4

ABR40414  
 ID ABR40414 standard; Protein: 263 AA.

XX ABR40414;

DT 13-JUN-2003 (first entry)

XX Human secreted protein #SEQ ID 164.

KW Human; secreted protein: anti-HIV; nootropic; neuroprotective;  
 KW antiangiinal; immunosuppressive; immunomodulator; cytostatic; cardiac;  
 KW hepatotropic; antiinflammatory; antiallergic; antidiabetic;  
 KW gastrointestinal; antiinfertility; nephrotropic; vinuclid; hypotensive;  
 KW vasotrophic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
 KW antiaesthetic; antiparasitic; cerebroprotective; antibacterial;  
 KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
 KW hyperproliferative disorder; leukaemia; autoimmune disorder;  
 KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
 KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
 KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
 KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
 KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
 KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
 KW respiratory disorder; infectious disease; chromosome identification;  
 KW food additive; nutrition.

XX Homo sapiens.

XX W0200268628-A1.

XX 06-SEP-2002.

XX

PF 21-FEB-2002; 2002MO-US05301.

XX 23-FEB-2001; 2001US-270625P.

PR 12-JUL-2001; 2001US-304417P.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Komatsoulis G, Baker KP, Fisceella M, Moore PA, Wei P;

PI Duan DR, Shi Y, Gupta R;

XX WPI: 2002-750417/81.

DR N-PSDB; AB282469.

XX New human secreted proteins and nucleic acids, useful for preventing,

PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune

PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,

PT obesity or cirrhosis

XX Claim 11; Page 755; 873pp; English.

XX The invention relates to novel human secreted proteins and the genes  
 CC encoding them. Genes and proteins of the invention may be useful for  
 CC preventing, treating or ameliorating medical conditions e.g. by protein  
 CC or gene therapy. These conditions include cancer and hyperproliferative  
 CC disorders, immune cell proliferative disorders (e.g. leukemia),  
 CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
 CC infertility, placental and uterine disorders (e.g. endometriosis),  
 CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
 CC disease), wound healing, gastrointestinal system disorders, particularly  
 CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
 CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
 CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
 CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
 CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
 CC nucleic acids are also useful for chromosome identification, radiation  
 CC hybrid mapping or long-range restriction mapping. The polypeptide,  
 CC polynucleotide, agonist or antagonist may also be used as a food additive  
 CC or preservative to increase or decrease storage capabilities, fat content  
 CC or other nutritional components. The sequences given in records  
 CC ABR40409-ABR40590 and AB282464-AB282611 represent human secreted proteins  
 CC and the genes encoding them.

XX Sequence 263 AA;

Query Match 100.0%; Score 1393; DB 23; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-120;  
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTAWQAFVSNMMLAEAYSGGCFWMDNGLYREDQTSAPGRLCLNMLDAOSGLASAP 60  
 Db 1 MLTAWQAFVSNMMLAEAYSGGCFWMDNGLYREDQTSAPGRLCLNMLDAOSGLASAP 60  
 QY 61 VSGAGNHSYCRNPDEDPGRWCYVSGAGVPEKRPCEDLPCPTTSQALPAFTTEIOEAS 120  
 Db 61 VSGAGNHSYCRNPDEDPGRWCYVSGAGVPEKRPCEDLPCPTTSQALPAFTTEIOEAS 120  
 QY 121 EGGADDEVQVAPAPNALPARSEAAAVQVIGISQRYMNSKEKKDLGTGLGVIGITMVI 180  
 Db 121 EGGADDEVQVAPAPNALPARSEAAAVQVIGISQRYMNSKEKKDLGTGLGVIGITMVI 180  
 QY 181 IIAIGAGIILIGYSYKRGKDKLKEQHDQVCEREMQRTLP.SAFTNPCEIYDEKTVVHT 240  
 Db 181 IIAIGAGIILIGYSYKRGKDKLKEQHDQVCEREMQRTLP.SAFTNPCEIYDEKTVVHT 240  
 QY 241 SQTPVDPQEGSTPLMGAGTPGA 263  
 Db 241 SQTPVDPQEGSTPLMGAGTPGA 263

#### RESULT 5

AAU86149  
 ID AAU86149 standard; Protein: 263 AA.

XX

```

AC AA086149;
XX
XX 15-JUL-2002 (first entry)
XX
XX Human PRO264 polypeptide.
XX
XX Human; PRO; benign tumour; malignant tumour; lymphoid malignancy;
XX Leukemia; neuronal disorder; stromal disorder; blastocoeleic disorder;
XX inflammatory disorder; immune disorder; angiogenic disorder;
XX cytostatic; neuroprotective.
OS Homo sapiens.
XX
XX WO200153486-A1.
XX
XX 26-JUL-2001.
XX
XX 11-FEB-2000; 2000WO-US03565.
XX
XX 08-MAR-1999; 99WO-US05028.
XX 11-MAR-1999; 99US-123972P.
XX 11-MAY-1999; 99US-133459P.
XX 02-JUN-1999; 99WO-US12252.
XX 22-JUN-1999; 99US-140650P.
XX 22-JUN-1999; 99US-140653P.
XX 20-JUL-1999; 99US-144758P.
XX 26-JUL-1999; 99US-145698P.
XX 28-JUL-1999; 99US-146222P.
XX 17-AUG-1999; 99US-149395P.
XX 31-AUG-1999; 99US-151689P.
XX 01-SEP-1999; 99WO-US20111.
XX 15-SEP-1999; 99WO-US21090.
XX 30-NOV-1999; 99WO-US28313.
XX 01-DEC-1999; 99WO-US28301.
XX 01-DEC-1999; 99WO-US28634.
XX 05-JAN-2000; 2000WO-US00219.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Hillan KJ;
XX Marsters SA, Pan J, Pilti RM, Roy MA, Smith V, Stone DM;
XX Watanabe CK, Wood WI;
XX
XX MPI; 2002-205567/26.
XX
XX N-PSDB; ABK40275.
XX
XX Thirty five nucleic acids encoding PRO polypeptides, useful for
XX treating benign or malignant tumours, leukemias and lymphoid
XX malignancies, inflammatory, angiogenic and immunologic disorders -
XX
XX Claim 61; Fig 44; 302pp; English.
XX
XX The present invention relates to the isolation of novel human PRO
XX polypeptides and the polynucleotide sequences encoding them. The
XX PRO polypeptides, agonists, antagonists or anti-PRO antibodies are
XX useful for treating benign or malignant tumours (e.g. renal, kidney,
XX bladder, breast, etc), leukemias and lymphoid malignancies, other
XX disorders such as neuronal, glial, astrocytal, hypothalamic, glandular,
XX macrophagal, stromal and blastocoeleic disorders, inflammatory, immune
XX and angiogenic disorders. The polynucleotide sequences are also
XX useful in gene therapy. AA086128-AA086162 represent the human PRO
XX polypeptides of the invention.
XX
XX Sequence 263 AA;
XX
XX Query Match 99.8%; Score 1390; DB 23; Length 263;
XX Best Local Similarity 99.6%; Pred. No. 1.6e-119;
XX Matches 262; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MLAAWQAFIVSNMILABAYGSGGCFWNGHLYREDQTSAPGRCIMLWILDAOSGLASAP 60
XX 1 MLAAWQAFIVSNMILABAYGSGGCFWNGHLYREDQTSAPGRCIMLWILDAOSGLASAP 60

```

```

QY VSGAGNHSYCRNPDEDPGRGCMVYSGAGVPEKRCEDLRCPETTSQALPAFTTEIOAS 120
DB 61 VSGAGNHSYCRNPDEDPGRGCMVYSGAGVPEKRCEDLRCPETTSQALPAFTTEIOAS 120
QY 121 BGPQADVEYQVAPANALPARSEAAAVQVIGISQFRVNSKXKDLGTLGYLGTMMVY 180
DB 121 BGPQADVEYQVAPANALPARSEAAAVQVIGISQFRVNSKXKDLGTLGYLGTMMVY 180
QY 181 IIAIAGIILIGYSYRGKDLKEQHDQKVCEREMRITLPLSAFTNPCEIYDEKTVVHT 240
DB 181 IIAIAGIILIGYSYRGKDLKEQHDQKVCEREMRITLPLSAFTNPCEIYDEKTVVHT 240
QY 241 SQTVPDQEGSTRPLMGQAGTPGA 263
DB 241 SQTVPDQEGSTRPLMGQAGTPGA 263

```

## RESULT 6

```

AAB43237
ID AAB43237 standard. Protein, 263 AA.
AC AAB43237;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF3001 polypeptide sequence SEQ ID NO:6002.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnery; antipapillary; antiparkinsonian; noctropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antineoplastic;
XX antiviral; antifungal; antihuman; antihuman; antihuman;
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Leach M;
XX
XX MPI; 2000-602362/57.
XX N-PSDB; AAC77446.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 5181-5182; 5507pp; English.
XX
XX AAC7446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antipapillary; antiparkinsonian; noctropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

```

CC immunostimulant; cardiant; thrombolytic; coagulant; vasoactive;  
 CC anti-diabetic; hypotensive; dermatological; immunosuppressive;  
 CC anti-inflammatory; antibacterial; antiviral; antifungal; antineoplastic;  
 CC antihypertensive; antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy.  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, anti-inflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 263 AA;

Query Match 99.4%; Score 1384; DB 21; Length 263;  
 Best Local Similarity 99.2%; Pred. No. 5.6e-119;  
 Matches 261; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLAAWQAFVSNMMLAEAYSGGCGFWDNGHLYREDQTSAPGLRCINMLDAOSGLASAP 60  
 Db 1 MLAAWQAFVSNMMLAEAYSGGCGFWDNGHLYREDQTSAPGLRCINMLDAOSGLASAP 60  
 QY 61 VSGAGNHSYCRNPDEDPGRWCYVSGAGVPEKPCEDLRCPETTSQALPAFTTEIOEAS 120  
 Db 61 VSGAGNHSYCRNPDEDPGRWCYVSGAGVPEKPCEDLRCPETTSQALPAFTTEIOEAS 120  
 QY 121 EGGADAEVQVFAFANALPARSEAAAVQVIGISQVRMNSKEKKDGLGLGVIGITMMVI 180  
 Db 121 EGGADAEVQVFAFANALPARSEAAAVQVIGISQVRMNSKEKKDGLGLGVIGITMMVI 180  
 QY 181 IIAAGAILIGYSYKRGKDLKEQHDQVCEREMQRIPLPSAFNPTCEIYDEKTVVHT 240  
 Db 181 IIAAGAILIGYSYKRGKDLKEQHDQVCEREMQRIPLPSAFNPTCEIYDEKTVVHT 240  
 QY 241 SQTPVDPQEGSTPLMGAGTPGA 263  
 Db 241 SQTPVDPQEGSTPLMGAGTPGA 263

RESULT 7  
 AAM93748  
 ID AAM93748 standard; Protein; 263 AA.

XX AC AAM93748;  
 XX DT 06-NOV-2001 (first entry)  
 XX DE Human polypeptide, SEQ ID NO: 3727.  
 XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.  
 XX OS Homo sapiens.  
 XX PN EP1130094-A2.  
 XX PD 05-SEP-2001.  
 XX PF 07-JUL-2000; 2000EP-011089.  
 XX PR 08-JUL-1999; 99JP-0194486.  
 XX PR 11-JAN-2000; 2000JP-018774.  
 XX PR 02-MAY-2000; 2000JP-0183765.  
 XX PA (HELI-) HELIX RES INST.  
 XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,  
 XX PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR WPI; 2001-524255/58.  
 DR N-PSDB; AAK94700.  
 XX 830 Primers useful for synthesizing full length cDNA clones and their  
 PT use in genetic manipulation -  
 PS Claim 8; SEQ ID NO 3727; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been  
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
 CC molecules have been determined. Primers for synthesizing the full length  
 CC cDNA are useful for clarifying the function of the protein encoded by  
 CC the cDNA. The full length clones were obtained by construction of full  
 CC length enriched cDNA libraries that were synthesised by the oligo-capping  
 CC method. The primers enable the production of the full length cDNA easily  
 CC without any special methods. The present sequence is a polypeptide  
 CC encoded by a full length human cDNA of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 263 AA;

Query Match 99.0%; Score 1379; DB 22; Length 263;  
 Best Local Similarity 99.2%; Pred. No. 1.6e-118;  
 Matches 261; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLAAWQAFVSNMMLAEAYSGGCGFWDNGHLYREDQTSAPGLRCINMLDAOSGLASAP 60  
 Db 1 MLAAWQAFVSNMMLAEAYSGGCGFWDNGHLYREDQTSAPGLRCINMLDAOSGLASAP 60  
 QY 61 VSGAGNHSYCRNPDEDPGRWCYVSGAGVPEKPCEDLRCPETTSQALPAFTTEIOEAS 120  
 Db 61 VSGAGNHSYCRNPDEDPGRWCYVSGAGVPEKPCEDLRCPETTSQALPAFTTEIOEAS 120  
 QY 121 EGGADAEVQVFAFANALPARSEAAAVQVIGISQVRMNSKEKKDGLGLGVIGITMMVI 180  
 Db 121 EGGADAEVQVFAFANALPARSEAAAVQVIGISQVRMNSKEKKDGLGLGVIGITMMVI 180  
 QY 181 IIAAGAILIGYSYKRGKDLKEQHDQVCEREMQRIPLPSAFNPTCEIYDEKTVVHT 240  
 Db 181 IIAAGAILIGYSYKRGKDLKEQHDQVCEREMQRIPLPSAFNPTCEIYDEKTVVHT 240  
 QY 241 SQTPVDPQEGSTPLMGAGTPGA 263  
 Db 241 SQTPVDPQEGSTPLMGAGTPGA 263

RESULT 8  
 AAY05220  
 ID AAY05220 standard; Protein; 286 AA.

XX AC AAY05220;  
 XX DT 17-JUN-1999 (first entry)  
 XX DE Kringiel protein sequence.  
 XX KW Kringiel; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
 KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 KW myocardial infarction; hypotension; hypertension; allergy; infection;  
 KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 KW male pattern baldness.  
 XX OS Homo sapiens.  
 XX PN WO9911788-A1.  
 XX PD 11-MAR-1999.



Db 1 MLAWQAFVLSNMLLAEGVSGGCFWMDGHLVREDQTSBPAPRLCLNMLDAQSGLASAP 60  
 QY 61 VSGAGNHSYCRNPDEDPGRCWCYVSGAGVPEKRPCEDJRCPTTSQALPAFTTEIQEAS 120  
 Db 61 VSGAGNHSYCRNPDEDPGRCWCYVSGAGVPEKRPCEDJRCPTTSQALPAFTTEIQEAS 120  
 QY 121 EGGGADEVQVFAPANALPARSEAAV 146  
 Db 121 EGGGADEVQVFAPATPPLGSAAXAV 146  
 RESULT 10  
 ABR40561  
 ID ABR40561 standard; Protein: 146 AA.  
 AC ABR40561;  
 XX 13-JUN-2003 (first entry)  
 DE Human secreted protein #SEQ ID 311.  
 XX  
 KW Human; secreted protein; anti-HIV; neurotropic; neuroprotective;  
 KW anti-angiogenic; immunosuppressive; immunomodulator; cytoprotective; cardiac;  
 KW hepatocellular; antiinflammatory; antiallergic; antidiabetic;  
 KW gastroenteric; antiparasitic; nephrotropic; vitruvian; hypotensive;  
 KW vasotrophic; dermatological; osteopathic; antiallergic; antiparkinsonian;  
 KW antiaesthetic; antiparasitic; cerebroprotective; antibacterial;  
 KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
 KW hyperproliferative disorder; leukaemia; autoimmune disorder;  
 KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
 KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
 KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
 KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
 KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
 KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
 KW respiratory disorder; infectious disease; chromosome identification;  
 KW food additive; nutrition.  
 XX  
 OS Homo sapiens.  
 XX  
 WN WO200268628-A1.  
 XX  
 PD 06-SEP-2002.  
 XX  
 PE 21-FEB-2002; 2002WO-US05301.  
 XX  
 PR 23-FEB-2001; 2001US-270625P.  
 PR 12-JUN-2001; 2001US-304417P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Komatsoulis G, Baker KP, Fiscella M, Moore PA, Wei P;  
 PI Duan DR, Shi Y, Gupta R;  
 XX  
 DR WPI; 2002-750417/81.  
 XX  
 PT New human secreted proteins and nucleic acids, useful for preventing,  
 PT treating, or ameliorating a medical condition, e.g. cancer, autoimmune  
 PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
 PT obesity or cirrhosis -  
 XX  
 PS Claim 11; Page 850; 873pp; English.  
 XX  
 CC The invention relates to novel human secreted proteins and the genes  
 CC encoding them. Genes and proteins of the invention may be useful for  
 CC preventing, treating or ameliorating medical conditions e.g. by protein  
 CC or gene therapy. These conditions include cancer and hyperproliferative  
 CC disorders, immune cell proliferative disorders (e.g. leukaemia),  
 CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
 CC infertility, placental and uterine disorders (e.g. endometriosis),  
 CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
 CC disease), wound healing, gastrointestinal system disorders, particularly

CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
 CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
 CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
 CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
 CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
 CC nucleic acids are also useful for chromosome identification, radiation  
 CC hybrid mapping or long-range restriction mapping. The polypeptide,  
 CC polynucleotide, agonist or antagonist may also be used as a food additive  
 CC or preservative to increase or decrease storage capabilities, fat content  
 CC or other nutritional components. The sequences given in records  
 CC ABR40409-ABR40590 and ABR282464-ABR282611 represent human secreted proteins  
 CC and the genes encoding them.  
 XX  
 SQ Sequence 146 AA;  
 XX  
 Query Match 54.0%; Score 752; DB 23; Length 146;  
 Best Local Similarity 94.5%; Pred. No. 3.4e-61;  
 Matches 138; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 MLAWQAFVLSNMLLAEGVSGGCFWMDGHLVREDQTSBPAPRLCLNMLDAQSGLASAP 60  
 Db 1 MLAWQAFVLSNMLLAEGVSGGCFWMDGHLVREDQTSBPAPRLCLNMLDAQSGLASAP 60  
 QY 61 VSGAGNHSYCRNPDEDPGRCWCYVSGAGVPEKRPCEDJRCPTTSQALPAFTTEIQEAS 120  
 Db 61 VSGAGNHSYCRNPDEDPGRCWCYVSGAGVPEKRPCEDJRCPTTSQALPAFTTEIQEAS 120  
 QY 121 EGGGADEVQVFAPANALPARSEAAV 146  
 Db 121 EGGGADEVQVFAPATPPLGSAAXAV 146  
 RESULT 11  
 ABR52752  
 ID ABR52752 standard; Peptide; 66 AA.  
 XX  
 AC ABR52752;  
 XX  
 DT 25-FEB-2003 (first entry)  
 XX  
 DE Human liver peptide, SEQ ID No 31400.  
 XX  
 KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157273-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PE 30-JAN-2001; 2001WO-US00664.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488898/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analysing gene expression in human adult liver -  
 XX  
 PS Claim 27; SEQ ID No 31400; 658pp; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe (SENp) (I) for



CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult  
CC liver. (I) may be used for predicting, measuring and displaying gene  
CC expression in samples derived from human adult liver. The genes  
CC identified may be involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
CC is associated with coronary heart disease. ABG47348-ABG59930 represent  
CC human liver single exon encoded peptides of the invention.  
CC Note: The sequence information for this patent does not appear in the  
CC printed specification but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
CC  
SQ Sequence 66 AA;  
Query Match 23.1%; Score 322; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 4.2e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 104 TTGALPFTTEIQASEGPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 163  
DB 1 TTGALPFTTEIQASEGPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
OY 164 KDLGTL 169  
DB 61 KDLGTL 66  
RESULT 12  
ABB37905  
ID ABB37905 standard; Peptide; 66 AA.  
XX  
AC ABB37905;  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #5411 encoded by human foetal liver single exon probe.  
XX  
KM Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
XX  
XX analyzing gene expression in human foetal liver -  
XX  
PS Claim 27; SEQ ID NO 30540; 639pp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a peptide encoded by a single exon

CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
CC  
SQ Sequence 66 AA;  
Query Match 23.1%; Score 322; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 4.2e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 104 TTGALPFTTEIQASEGPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 163  
DB 1 TTGALPFTTEIQASEGPGADEVOVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
OY 164 KDLGTL 169  
DB 61 KDLGTL 66  
RESULT 13  
ABB23159  
ID ABB23159 standard; Protein; 66 AA.  
XX  
AC ABB23159;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #5158 encoded by probe for measuring heart cell gene expression.  
XX  
KM Human; gene expression; heart; microarray; vascular system;  
XX  
KM cardiovascular disease; hypertension; cardiac arrhythmia;  
XX  
KM congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX  
XX hearts -  
XX  
PS Claim 15; SEQ ID NO 24929; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 66 AA;  
SQ Query Match 23.1%; Score 322; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 4.2e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TTSQLPFTTEIOEASBGDADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 163  
DB 1 TTSQLPFTTEIOEASBGDADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
QY 164 KDLGTL 169  
DB 61 KDLGTL 66

RESULT 14  
AAM58537  
ID AAM58537 standard; Protein: 66 AA.  
XX  
AC AAM58537;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30642.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN W0200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001WO-US00667.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
PS Example 4; SEQ ID NO: 30642; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
SQ Sequence 66 AA;

Query Match 23.1%; Score 322; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 4.2e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TTSQLPFTTEIOEASBGDADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 163  
DB 1 TTSQLPFTTEIOEASBGDADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60

DB 1 TTSQLPFTTEIOEASBGDADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
QY 164 KDLGTL 169  
DB 61 KDLGTL 66

RESULT 15  
AAM71037  
ID AAM71037 standard; Protein: 66 AA.  
XX  
AC AAM71037;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31343.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KM microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN W0200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488900/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
PS Example 4; SEQ ID NO: 31343; 658bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.  
XX  
SQ Sequence 66 AA;

Query Match 23.1%; Score 322; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 4.2e-22;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TTSQLPFTTEIOEASBGDADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 163  
DB 1 TTSQLPFTTEIOEASBGDADEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
QY 164 KDLGTL 169  
DB 61 KDLGTL 66

Search completed: November 25, 2003, 13:24:40  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:33:18 ; Search time 30 Seconds

(without alignments)  
1616.953 Million cell updates/sec

Title: US-10-057-951-2

Perfect score: 1393  
Sequence: 1 MLAMVQAFIVSNMLAEAY.....PVDFQEGSTPLMGAGTPGA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673684 segs, 184443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1393	100.0	263	9	US-09-084-491A-2
2	1393	100.0	263	14	US-10-102-704-2
3	1393	100.0	263	14	US-10-057-951-2
4	1390	99.8	263	12	US-10-210-951-44
5	1390	99.8	263	12	US-10-211-884-44
6	1322	23.1	66	9	US-09-864-761-38457
7	154.5	11.1	527	11	US-09-987-457-18
8	154.5	11.1	527	11	US-09-987-455-19
9	154.5	11.1	562	9	US-09-969-271-7
10	154.5	11.1	562	10	US-09-974-298-145
11	154.5	11.1	562	12	US-10-443-701-4
12	154.5	11.1	562	15	US-10-193-656-8
13	146.5	10.5	655	15	US-10-172-712-28
14	142	10.2	160	9	US-09-761-120-35
15	142	10.2	160	10	US-09-335-325-35

16	142	10.2	160	15	US-10-131-241-35	Sequence 35, Appl
17	142	10.2	250	9	US-09-761-120-30	Sequence 30, Appl
18	142	10.2	250	10	US-09-335-325-30	Sequence 30, Appl
19	142	10.2	250	15	US-10-131-241-30	Sequence 30, Appl
20	142	10.2	260	15	US-10-131-241-61	Sequence 61, Appl
21	142	10.2	339	9	US-09-788-142-3	Sequence 3, Appl
22	142	10.2	339	9	US-09-761-120-3	Sequence 3, Appl
23	142	10.2	339	10	US-09-335-325-3	Sequence 3, Appl
24	142	10.2	339	15	US-10-131-241-3	Sequence 40, Appl
25	142	10.2	352	9	US-09-761-120-40	Sequence 40, Appl
26	142	10.2	352	10	US-09-335-325-40	Sequence 40, Appl
27	142	10.2	352	15	US-10-131-241-40	Sequence 40, Appl
28	142	10.2	363	12	US-10-292-418-11	Sequence 11, Appl
29	142	10.2	368	9	US-09-761-120-42	Sequence 42, Appl
30	142	10.2	378	9	US-09-873-676-1	Sequence 1, Appl
31	142	10.2	378	10	US-09-335-325-42	Sequence 42, Appl
32	142	10.2	378	15	US-10-131-241-42	Sequence 42, Appl
33	142	10.2	391	15	US-10-304-287-7	Sequence 7, Appl
34	142	10.2	394	15	US-10-304-287-8	Sequence 8, Appl
35	142	10.2	458	9	US-09-946-893-4	Sequence 4, Appl
36	142	10.2	569	9	US-09-946-893-5	Sequence 5, Appl
37	142	10.2	571	9	US-09-946-893-8	Sequence 8, Appl
38	142	10.2	576	9	US-09-946-893-6	Sequence 6, Appl
39	142	10.2	791	10	US-09-967-386-1	Sequence 1, Appl
40	142	10.2	791	15	US-10-304-287-1	Sequence 1, Appl
41	142	10.2	810	9	US-09-946-893-2	Sequence 2, Appl
42	142	10.2	810	12	US-10-237-144-1	Sequence 1, Appl
43	142	10.2	810	15	US-10-193-656-2	Sequence 2, Appl
44	138	9.9	79	9	US-09-753-064-2	Sequence 2, Appl
45	138	9.9	79	9	US-09-761-120-8	Sequence 8, Appl

## ALIGNMENTS

RESULT 1  
US-09-084-491A-2  
; Sequence 2, Application US/09084491A  
; Patent No. US20020061576A1  
; GENERAL INFORMATION:  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: BENNER, REINHARD  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR-LIKE PROTEASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
; STREET: 9410 KEY WEST AVENUE  
; CITY: ROCKVILLE  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/084,491A  
; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROOKES, ANDERS A.  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF78  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 263 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
US-09-084-491A-2

Query Match 100.0%; Score 1393; DB 9; Length 263;  
Best Local Similarity 100.0%; Pred. No. 8.9e-126;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLAMVQAFVLSNMLLAEAYSGGCFWMDNGHLYREDOTSPAPGLRCLNMLDAOSGLASAP 60
DB 1 MLAMVQAFVLSNMLLAEAYSGGCFWMDNGHLYREDOTSPAPGLRCLNMLDAOSGLASAP 60
QY 61 VSGAGNHSYCRNPDEDPDRGPMWCYSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120
DB 61 VSGAGNHSYCRNPDEDPDRGPMWCYSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120
QY 121 EGPDADEVQVFAPANALPARSEAAAVQPIGISOVRNMSKEKKDGLTGYVLGITMAYI 180
DB 121 EGPDADEVQVFAPANALPARSEAAAVQPIGISOVRNMSKEKKDGLTGYVLGITMAYI 180
QY 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTVVHT 240
DB 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTVVHT 240
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263

```

## RESULT 2

US-10-102-704-2

Sequence 2, Application US/10102704  
Publication No. US20020164768A1

GENERAL INFORMATION:

APPLICANT: Moore et al.  
TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protein  
FILE REFERENCE: P378C1  
CURRENT APPLICATION NUMBER: US/10/102,704  
CURRENT FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: 09/084,491  
PRIOR FILING DATE: 1998-05-27  
PRIOR APPLICATION NUMBER: 60/048,000  
PRIOR FILING DATE: 1997-05-28  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 263  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-102-704-2

Query Match 100.0%; Score 1393; DB 14; Length 263;  
Best Local Similarity 100.0%; Pred. No. 8.9e-126;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLAMVQAFVLSNMLLAEAYSGGCFWMDNGHLYREDOTSPAPGLRCLNMLDAOSGLASAP 60
DB 1 MLAMVQAFVLSNMLLAEAYSGGCFWMDNGHLYREDOTSPAPGLRCLNMLDAOSGLASAP 60
QY 61 VSGAGNHSYCRNPDEDPDRGPMWCYSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120
DB 61 VSGAGNHSYCRNPDEDPDRGPMWCYSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120
QY 121 EGPDADEVQVFAPANALPARSEAAAVQPIGISOVRNMSKEKKDGLTGYVLGITMAYI 180
DB 121 EGPDADEVQVFAPANALPARSEAAAVQPIGISOVRNMSKEKKDGLTGYVLGITMAYI 180
QY 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTVVHT 240
DB 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTVVHT 240
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263

```

## RESULT 3

US-10-057-951-2

Sequence 2, Application US/10057951  
Publication No. US2002017713A1

GENERAL INFORMATION:

APPLICANT: Moore et al.  
TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease  
FILE REFERENCE: P378P1  
CURRENT APPLICATION NUMBER: US/10/057,951  
CURRENT FILING DATE: 2002-01-29  
PRIOR APPLICATION NUMBER: US 09/411,977  
PRIOR FILING DATE: 1999-10-04  
PRIOR APPLICATION NUMBER: US 09/084,491  
PRIOR FILING DATE: 1998-05-27  
PRIOR APPLICATION NUMBER: US 60/048,000  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: PatentIn Ver. 3.1  
SEQ ID NO 2  
LENGTH: 263  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-057-951-2

Query Match 100.0%; Score 1393; DB 14; Length 263;  
Best Local Similarity 100.0%; Pred. No. 8.9e-126;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLAMVQAFVLSNMLLAEAYSGGCFWMDNGHLYREDOTSPAPGLRCLNMLDAOSGLASAP 60
DB 1 MLAMVQAFVLSNMLLAEAYSGGCFWMDNGHLYREDOTSPAPGLRCLNMLDAOSGLASAP 60
QY 61 VSGAGNHSYCRNPDEDPDRGPMWCYSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120
DB 61 VSGAGNHSYCRNPDEDPDRGPMWCYSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120
QY 121 EGPDADEVQVFAPANALPARSEAAAVQPIGISOVRNMSKEKKDGLTGYVLGITMAYI 180
DB 121 EGPDADEVQVFAPANALPARSEAAAVQPIGISOVRNMSKEKKDGLTGYVLGITMAYI 180
QY 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTVVHT 240
DB 181 IIAIGAGIILGYSKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTVVHT 240
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263

```

## RESULT 4

US-10-210-951-44

Sequence 44, Application US/10210951  
Publication No. US20030170228A1

GENERAL INFORMATION:

APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Marcsters, Scott A.  
APPLICANT: Pan, James  
APPLICANT: Picti, Robert M.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Smith, Victoria  
APPLICANT: Stone, Donna M.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR  
FILE REFERENCE: P293IRIC1  
CURRENT APPLICATION NUMBER: US/10/210,951  
CURRENT FILING DATE: 2002-08-02

```

; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063445
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SEQ ID NO 44
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-210-951-44
```

Query Match 99.8%; Score 1390; DB 12; Length 263;

Best Local Similarity 99.6%; Pred. No. 1.7e-125; Matches 262; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLAAVQAFIVSNMLLAEAAYGSGGCFWMDNGHLYREDQTSAPAGLRCLNMLDAQSGIASAP 60
DB 1 MLAAVQAFIVSNMLLAEAAYGSGGCFWMDNGHLYREDQTSAPAGLRCLNMLDAQSGIASAP 60
QY 61 VSGAGNHSCRNPDDEDPGRCWCYSGEAGVPEKRCEDIRCEPETSQALPAFTTEIOEAS 120
DB 61 VSGAGNHSCRNPDDEDPGRCWCYSGEAGVPEKRCEDIRCEPETSQALPAFTTEIOEAS 120
QY 121 EBPGEDEVQFAPANALPARSEAAVOPVIGISQVRNMSKEKDLGTGLGYLGITMVI 180
DB 121 EBPGEDEVQFAPANALPARSEAAVOPVIGISQVRNMSKEKDLGTGLGYLGITMVI 180
QY 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTPLSAFTNPTCEIVDEKTVVHT 240
DB 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTPLSAFTNPTCEIVDEKTVVHT 240
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263
```

RESULT 5  
US-10-211-884-44

; Sequence 44, Application US/10211884  
; Publication No. US20030175900A1

; GENERAL INFORMATION:

```

; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Masters, Scot A.
; APPLICANT: Pan, James
; APPLICANT: Pilti, Robert M.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stone, Donna M.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2931R1C1
```

```

; CURRENT APPLICATION NUMBER: US/10/211,884
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/014699
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: 60/026943
; PRIOR FILING DATE: 1996-09-23
; PRIOR APPLICATION NUMBER: 60/059121
; PRIOR FILING DATE: 1997-07-17
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/062037
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 258
; SEQ ID NO 44
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-884-44
```

Query Match 99.8%; Score 1390; DB 12; Length 263;

Best Local Similarity 99.6%; Pred. No. 1.7e-125; Matches 262; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLAAVQAFIVSNMLLAEAAYGSGGCFWMDNGHLYREDQTSAPAGLRCLNMLDAQSGIASAP 60
DB 1 MLAAVQAFIVSNMLLAEAAYGSGGCFWMDNGHLYREDQTSAPAGLRCLNMLDAQSGIASAP 60
QY 61 VSGAGNHSCRNPDDEDPGRCWCYSGEAGVPEKRCEDIRCEPETSQALPAFTTEIOEAS 120
DB 61 VSGAGNHSCRNPDDEDPGRCWCYSGEAGVPEKRCEDIRCEPETSQALPAFTTEIOEAS 120
QY 121 EBPGEDEVQFAPANALPARSEAAVOPVIGISQVRNMSKEKDLGTGLGYLGITMVI 180
DB 121 EBPGEDEVQFAPANALPARSEAAVOPVIGISQVRNMSKEKDLGTGLGYLGITMVI 180
QY 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTPLSAFTNPTCEIVDEKTVVHT 240
DB 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTPLSAFTNPTCEIVDEKTVVHT 240
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263
```

RESULT 6  
US-09-864-761-38457

; Sequence 38457, Application US/09864761  
; Patent No. US20020048763A1

; GENERAL INFORMATION:

```

; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
```

```

/ APPLICANT: Goetz, Friedrich
/ APPLICANT: Werner, Rolf-Guenther
/ TITLE OF INVENTION: Methods for Large Scale Protein Production in Prokaryotes
/ FILE REFERENCE: 0652.2180001
/ CURRENT APPLICATION NUMBER: US/09/987,457
/ CURRENT FILING DATE: 2001-11-14
/ PRIOR APPLICATION NUMBER: 60/268,573
/ PRIOR FILING DATE: 2001-02-15
/ PRIOR APPLICATION NUMBER: GB 00 27 782.2
/ PRIOR FILING DATE: 2000-11-14
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 18
/ LENGTH: 527
/ TYPE: PRT
/ ORGANISM: Homo sapiens (tPA)
US-09-987-457-18

Query Match 11.1%; Score 154.5; DB 11; Length 527;
Best Local Similarity 39.6%; Pred. No. 3.4e-06;
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 25 CFWDNGHLYREDQTSAPGRLCLNWLDAQSLASAPV-----GAGNHSYCNPNDE 75
DB 92 CYEDDGISYFGTWSITESGAECTNW--NSSALAKPYSGRRPDAILRLGLGNHNYCRNPD 149
66 DPGPWCYVSGEAGVPEKRPCEDLRCPEPTS 106
150 DSK-PWCYVF-KAGKYSSEFCSTPACSGNS 178

RESULT 9
US-09-969-271-7

```



US-10-172-712-28  
Sequence 28, Application US/10172712  
Publication No. US20030125232A1  
GENERAL INFORMATION:  
APPLICANT: GRIFFIN, JOHN H.  
APPLICANT: GALE, ANDREW J.  
APPLICANT: GETZOFF, ELIZABETH D.  
APPLICANT: PELLEQUER, JEAN-LUC  
TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS  
FILE REFERENCE: 4198-8001US1  
CURRENT APPLICATION NUMBER: US/10/172,712  
CURRENT FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: 60/298,578  
PRIOR FILING DATE: 2001-06-14  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 28  
LENGTH: 655  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-172-712-28

	Query Match	Similarity	10.5%	Score	146.5	DB	15	Length	655
	Best local	Similarity	36.9%	Pred.	No.2.7e-05				
	Matches	Conservative	41	Mismatches	41	Indels	19	Gaps	6
QY	25	CFWMDGHLTYREDQTSPAPGLRCLTM-----LDASGLASPVSAGNHSYCNPDPD	76						
		:							
Dd	286	CFLGGGTGRGVASTSASGLSCLANMSDLIYLDELHVD--VGAALLGLGPAPHYKCNPNPD	344						
QY	77	PRGPCVCYSGEAGVP-----EKRPCEDRICPRTTSQALPAFTTEIOE-ASEG	122						
		:							
Dd	345	ER-PWCYVVKDSALSNFYCRLEACSEL-----TRVLSPLDLTLTPPPAPRG	390						

```

RESULT 14
US-09-761-120-35
; Sequence 35, Application US/09761120
; Patent No. US20020037847A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael
; APPLICANT: Folman, M. Judah
; TITLE OF INVENTION: Nucleic Acids Encoding Kringle 1-5 Region Fragments of Plasminogen
; FILE REFERENCE: 05940-0151 (43171-252068)
; CURRENT APPLICATION NUMBER: US/09/761,120
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 09/309,821
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 08/866,735
; PRIOR FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 35
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Kringle 1-2
US-09-761-120-35

```

```

Query Match          10.2%; Score 142; DB 9; Length 160;
Best Local Similarity 37.2%; Pred. No. 1,1e-05;
Matches      32; Conservative    8; Mismatches   34; Indels    12; Gaps     4;

QY      25 CFMDNGHLREDTSAFGRCALNWLDA-----OSGLASAPSGAGNHSCYNCPDEDPGG 79
         |||||
Db       1 CKTGANGNRKGMSTKRNGITCGKMSSTSHRFRSPATHPSEGL-EENYCNPNNDPPG 59

QY      80 PWCVYSGEAGVPEKR--PCEDLRCP 103
         |||||
Db       .60 PWCYTLD---PEKRYDYCDILECEE 81

```

```

1      RESULT 15
2      US-09-335-325-35
3      / Sequence 35, Application US/09335325
4      / Patent No. US20020164717A1
5      / GENERAL INFORMATION:
6      / APPLICANT: Folkman, M. Judah
7      / O'Reilly, Micheal
8      / Cao, Yihai
9      / Sim, B. Kim lee
10     / TITLE OF INVENTION: Angiotensin Fragments and Method of Use
11     / NUMBER OF SEQUENCES: 45
12     / CORRESPONDENCE ADDRESS:
13     / ADDRESSEE: Jones & Askew
14     / STREET: 191 Peachtree Street, 37th Floor
15     / CITY: Atlanta
16     / STATE: Georgia
17     / COUNTRY: U.S.
18     / ZIP: 30303-1769
19     / COMPUTER READABLE FORM:
20     / MEDIUM TYPE: Floppy disk
21     / COMPUTER: IBM PC compatible
22     / OPERATING SYSTEM: PC-DOS/MS-DOS
23     / SOFTWARE: Patentin Release #1.0, Version #1.30
24     / CURRENT APPLICATION DATA:
25     / APPLICATION NUMBER: US/09/335,325
26     / FILING DATE: 17-Jun-1999
27     / CLASSIFICATION: <Unknown>
28     / PRIOR APPLICATION DATA:
29     / APPLICATION NUMBER: US/08/612,788
30     / FILING DATE: <Unknown>
31     / ATTORNEY/AGENT INFORMATION:
32     / NAME: Warren, William L.
33     / REGISTRATION NUMBER: 36,714
34     / REFERENCE/DOCKET NUMBER: 05213-0126
35     / TELECOMMUNICATION INFORMATION:
36     / TELEPHONE: 404-818-3700
37     / TELEFAX: 404-818-3799
38     / INFORMATION FOR SEQ ID NO: 35:
39     / SEQUENCE CHARACTERISTICS:
40     / LENGTH: 160 amino acids
41     / TYPE: amino acid
42     / STRANDEDNESS: <Unknown>
43     / TOPOLOGY: linear
44     / MOLECULE TYPE: protein
45     / HYPOTHEICAL: NO
46     / ANTI-SENSE: NO
47     / FRAGMENT TYPE: N-terminal
48     / ORIGINAL SOURCE:
49     / ORGANISM: Homo sapiens
50     / IMMEDIATE SOURCE:
51     / CLONE: KI-2
52     / SEQUENCE DESCRIPTION: SEQ ID NO: 35:
53     /
54     / US-09-335-325-35

```

```

Query Match      10.2%; Score 142; DB 10; Length 160;
Best Local Similarity 37.2%; Pred. No. 1.1e-05;
Matches 32; Conservative 8; Mismatches 34; Indels 12; Gaps 4

QY      25 CFWDNHLTREDQTSFAPGLRCINLWDA-----QGSLASAPVSGAGNNHSCNPDPDPRG 79
          |   |||   |   :   |   :   |   :   |   :   |   :   |   :   |   :
Db       1 CKTGNGKNKRGMSKMGKITCGKNSSSTPHRPFSPATHPEGL-EENRYCNPNDPFG 59
          |   |||   |   :   |   :   |   :   |   :   |   :   |   :

QY      80 PWCYSGEAGVPBKR--PCEDLRCPF 103
          |||   :   |||   :   |||   :   |||   :   |||   :   |||   :
Db      60 PWCYTLD---PEKRYDYCDILECEE 81

```

Search completed: November 25, 2003, 13:38:47  
Job time : 31 secs

Search completed: November 25, 2003, 13:38:47  
Job time : 31 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:31:13 ; Search time 21 Seconds  
(without alignments)  
529.893 Million cell updates/sec

Title: US-10-057-951-2

Perfect score: 1393

Sequence: 1 MLAMVQAFVLSNMILAEAY.....PVDQEGSTPLMGQAGTPGA 263

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*\n2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*\n3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*\n4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*\n5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*\n6: /cgn2\_6/ptodata/1/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1393	100.0	263	4	US-09-411-977-2
2	163.5	11.7	472	2	US-08-811-949-63
3	154.5	11.1	437	2	US-08-811-949-49
4	154.5	11.1	437	2	US-08-811-949-51
5	154.5	11.1	437	2	US-08-811-949-55
6	154.5	11.1	437	2	US-08-811-949-57
7	154.5	11.1	527	2	US-07-609-510B-16
8	154.5	11.1	527	2	US-08-811-949-39
9	154.5	11.1	527	5	PCT-US91-01025A-2
10	154.5	11.1	527	6	5185259-8
11	154.5	11.1	527	6	5520913-1
12	154.5	11.1	546	6	5200340-6
13	154.5	11.1	562	2	US-08-811-949-43
14	154.5	11.1	562	2	US-08-560-098A-50
15	154.5	11.1	562	2	US-08-883-795A-38
16	154.5	11.1	562	6	5185259-3
17	154.5	11.1	562	6	5200340-2
18	154.5	11.1	562	6	5344773-2
19	148.5	10.7	83	2	US-08-811-949-2
20	146.5	10.5	655	1	US-08-148-910-12
21	146.5	10.5	655	1	US-08-448-937A-12
22	142	10.2	160	2	US-08-612-788-35
23	142	10.2	160	3	US-09-066-028-35
24	142	10.2	160	4	US-09-335-325-35
25	142	10.2	250	3	US-08-612-788-30
26	142	10.2	250	3	US-09-066-028-30
27	142	10.2	250	4	US-09-335-325-30

28	142	10.2	339	1	US-08-248-629A-3	Sequence 3, Appli
29	142	10.2	339	1	US-08-451-932-3	Sequence 3, Appli
30	142	10.2	339	1	US-08-452-260-3	Sequence 3, Appli
31	142	10.2	339	1	US-08-326-785-3	Sequence 3, Appli
32	142	10.2	339	2	US-08-612-788-3	Sequence 3, Appli
33	142	10.2	339	2	US-08-605-598B-3	Sequence 3, Appli
34	142	10.2	339	2	US-08-429-743-3	Sequence 3, Appli
35	142	10.2	339	2	US-08-866-733-3	Sequence 3, Appli
36	142	10.2	339	4	US-09-066-028-3	Sequence 3, Appli
37	142	10.2	339	4	US-09-335-325-3	Sequence 3, Appli
38	142	10.2	339	5	PCT-US95-05107-3	Sequence 3, Appli
39	142	10.2	352	2	US-08-612-788-40	Sequence 40, Appli
40	142	10.2	352	3	US-09-066-028-40	Sequence 40, Appli
41	142	10.2	352	4	US-09-335-325-40	Sequence 40, Appli
42	142	10.2	374	4	US-09-377-250-3	Sequence 3, Appli
43	142	10.2	378	4	US-09-377-250-2	Sequence 2, Appli
44	142	10.2	378	2	US-08-612-788-42	Sequence 42, Appli
45	142	10.2	378	3	US-09-066-028-42	Sequence 42, Appli

## ALIGNMENTS

RESULT 1  
US-09-411-977-2  
; Sequence 2, Application US/09411977  
; Patent No. 6372473  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Paul A.  
; APPLICANT: Ruben, Steven M.  
; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease  
; FILE REFERENCE: P37891  
; CURRENT APPLICATION NUMBER: US/09/411.977  
; EARLIER FILING DATE: 1999-10-04  
; EARLIER APPLICATION NUMBER: 09/084.491  
; EARLIER FILING DATE: 1998-05-27  
; EARLIER APPLICATION NUMBER: 60/048.000  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-411-977-2

Query Match 100.0%; Score 1393; DB 4; Length 263;  
Best Local Similarity 100.0%; Pred. No. 2.2e-139;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAMVQAFVLSNMILAEAYSGGCFPMNGHLIYREDQTS PAPGLRCLNWLDAQSLASAP 60  
1 MLAMVQAFVLSNMILAEAYSGGCFPMNGHLIYREDQTS PAPGLRCLNWLDAQSLASAP 60  
DB 1 VSGAGHSTCRPPDDPRGPMCVSGEAVPEKRPEDIRCEETTSQALPAFTTEIOEAS 120  
61 VSGAGHSTCRPPDDPRGPMCVSGEAVPEKRPEDIRCEETTSQALPAFTTEIOEAS 120  
QY 61 VSGAGHSTCRPPDDPRGPMCVSGEAVPEKRPEDIRCEETTSQALPAFTTEIOEAS 120  
DB 61 VSGAGHSTCRPPDDPRGPMCVSGEAVPEKRPEDIRCEETTSQALPAFTTEIOEAS 120  
QY 121 EGPQADEVQFPANALPARSEAAVQPIGTSORVRNMSKEKKDLGTLGYVLTMMYI 180  
121 EGPQADEVQFPANALPARSEAAVQPIGTSORVRNMSKEKKDLGTLGYVLTMMYI 180  
DB 121 EGPQADEVQFPANALPARSEAAVQPIGTSORVRNMSKEKKDLGTLGYVLTMMYI 180  
QY 181 IIAIGAGIILGYSYKRGKDLKEQHQKVCEREMORTTLPISAFNFTCEIVDEKTVVHT 240  
181 IIAIGAGIILGYSYKRGKDLKEQHQKVCEREMORTTLPISAFNFTCEIVDEKTVVHT 240  
DB 181 IIAIGAGIILGYSYKRGKDLKEQHQKVCEREMORTTLPISAFNFTCEIVDEKTVVHT 240  
QY 241 SQTPVDPOEGSTPLMGQAGTPGA 263  
241 SQTPVDPOEGSTPLMGQAGTPGA 263  
DB 241 SQTPVDPOEGSTPLMGQAGTPGA 263

RESULT 2

```

1 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
2 CITY: ARLINGTON
3 STATE: VA
4 COUNTRY: USA
5 ZIP: 22202
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: Patent Release #1.0, Version #1.30
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/811,949
14 FILING DATE: 05-MAR-1997
15 CLASSIFICATION: 435
16 ATTORNEY/AGENT INFORMATION:
17 NAME: OBLON, NORMAN F.
18 REGISTRATION NUMBER: 24,618
19 REFERENCE/DOCKET NUMBER: 18-966-0
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: 703-413-3000
22 TELEFAX: 703-413-2220
23 INFORMATION FOR SEQ ID NO: 49:
24 SEQUENCE CHARACTERISTICS:
25 LENGTH: 437 amino acids
26 TYPE: amino acid
27 TOPOLOGY: linear
28 MOLECULE TYPE: protein
29
30 US-08-811-949--49
31
32 Query Match 11.1%; Score 154.5; DB 2; Length 437;
33 Best Local Similarity 39.6%; Pred. No. 7.4e+08;
34 Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4
35
36 QY 25 CFWNGHLYREDQTSAPGLRCLMWLDAGSLASAPV-----GAGNHSYGRNPDE 75
37 | :|::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 Db 2 CYEGGISYRTGTWTASGABECTNW--NSSALAQKPYSGRRPDAILRLIGNHNYCRNPDR 59
39 | :|::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
40 QY 76 DRPGMVCVSGEAGVPEKRPCEDIRCPETTS 106
41 | :|::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
42 Db 60 DSK-PWCYVF-KAKGYSESEFCSTPACSGNS 88
43 | :|::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44
45 RESULT 4
46 US-08-811-949-51
47 Sequence 51, Application US/08811949
48 Patent No. 5840533
49 GENERAL INFORMATION:
50 APPLICANT: NIMWA, MINEO
51 APPLICANT: SAITO, YOSHIMASA
52 APPLICANT: SASAKI, HITOSHI
53 APPLICANT: HAYASHI, MASAKO
54 APPLICANT: NOTANI, JOUJI
55 APPLICANT: KOHAYASHI, MASAKAZU
56 TITLE OF INVENTION: TISSUE PLASMINGEN ACTIVATOR
57 NUMBER OF SEQUENCES: 67
58 CORRESPONDENCE ADDRESS:
59 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATIER & NEUSTADT,
60 ADDRESSEE: P.C.
61 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
62 CITY: ARLINGTON
63 STATE: VA
64 COUNTRY: USA
65 ZIP: 22202
66 COMPUTER READABLE FORM:
67 MEDIUM TYPE: Floppy disk
68 COMPUTER: IBM PC compatible
69 OPERATING SYSTEM: PC-DOS/MS-DOS
70 SOFTWARE: Patent Release #1.0, Version #1.30
71 CURRENT APPLICATION DATA:
72 APPLICATION NUMBER: US/08/811,949
73 FILING DATE: 05-MAR-1997
74 CLASSIFICATION: 435
75 ATTORNEY/AGENT INFORMATION:

```

11.18; Score 154.5; DB 2; Length 437;

; sequence 16, applica

to the service of the

; sequence 16, Application US/0/609310E

```

Patent No. 5326700
GENERAL INFORMATION:
APPLICANT: Berg et al
TITLE OF INVENTION: Method for Altering Post-Translational Processing of Tissue P
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN.
COUNTRY: U.S.A.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/609,510B
FILING DATE: 19901106
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 527 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
OS-07-609-510B-16

```

Query Match	11.1%;	Score 154.5;	DB 1;	Length 527;
Best Local Similarity	39.6%;	Pred. No. 9.8e-08;		
Matches 36;	Conservative 6;	Mismatches 36;	Indels 13;	Gaps 4;

QY 22 CFMDNHLKRPDDTSPAPRLCTNMDAOSGLASPVS-----GAGHSYCRNPDE 75  
Db 92 CYDDGSLGIRGTMTSLGSAECTNM--NSSLAQCPYGRRPDAIRLDGHNHCRNPDR 149  
QY 76 DPRGPWCYSGEAGVPEKRRCEDLRCPEFTTS 106  
Db 150 DSK-IPWCYVF-KAGKYSSEFCSTPAISEGNS 178

Db 150 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 178

RESULT 8  
 US-08-811-949-39  
 Sequence 39, Application US/08811949  
 Patent No. 5840533  
 GENERAL INFORMATION:  
 APPLICANT: NIMA, MINEO  
 APPLICANT: SAITO, YOSHIMASA  
 APPLICANT: SASAKI, HITOSHI  
 APPLICANT: HAYASHI, MASAKO  
 APPLICANT: NOTANI, JOUJI  
 APPLICANT: KOBAYASHI, MASAKAZU  
 TITLE OF INVENTION: TISSUE FLAMINGOGEN ACTIVATOR  
 NUMBER OF SEQUENCES: 67  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
 ADDRESSEE: P. C.  
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 CITY: ARLINGTON  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/811,949  
 FILING DATE: 05-MAR-1997  
 CLASSIFICATION: 435

```

/ ATTORNEY/AGENT INFORMATION:
/ NAME: OBLON, NORMAN F.
/ REGISTRATION NUMBER: 24,618
/ REFERENCE/DOCKET NUMBER: 18-966-0
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 703-413-3000
/ TELEFAX: 703-413-2220
/ INFORMATION FOR SEQ ID NO: 39:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 527 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-811-949-39

```

Query Match	Score	DB 2;	Length
11.1%;	154.5;	DB 2;	527;
Best Local Similarity	39.6%;	Best	98e-08;

Matches	36;	Conservative	6;	Mismatches	36;	Indels	13;	Gaps	4
---------	-----	--------------	----	------------	-----	--------	-----	------	---

Oy 25 C E F D N G H I Y R E D T S P A G L R C L M L D A O S G L A P A V S - - - - - G A C N H S Y C R N P D E 75  
Db 92 C Y E D O G I S Y R G T M S T A E S G A C E T W - - N S A L L A K P Y S G R R P D A T R L G L G H N H Y C R N P D R 149

```

Qy      76 DPGFMCYVSGEAGVPEKPCEDLRCPETTS 106
          | : |||| : ||
Db      150 DSK-FWCYVF-KAGKYSSEFCSTPACSEGN 178

```

RESULT 9  
PCT-US91-01025A-2

; Sequence 2, Application PC/TUS9101025A

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.  
TITLE OF INVENTION: Tissue Plasminogen Activator Having Fibrinolytic Activity

TITLE OF INVENTION: Specific Properties

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: Calif

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

```

;
;
COMPIER:  IBM PC compatible
OPERATING SYSTEM:  PC-DOS/MS-DOS

```

**SOFTWARE:** patin (Genentech)

```
; CURRENT APPLICATION DATA:
; DATE/TIME: 01/01/81 09:00
```

APPLICATION NUMBER: PCI/0591/010235A  
FILING DATE: 19910214

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 01/4  
FILING DATE: 1 March 199

ATTORNEY/AGENT INFORMATION

NAME: Hasak, Janet E.

REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 454B3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/266-18

TELEFAX: 415/952-9881  
TELEF: 910/371-7168

INFORMATION FOR SEO ID NO:

## SEQUENCE CHARACTERISTICS

LENGTH: 527 amino acids

TOPOLGY: linear

PCT-US91-01025A-2

Query March	11.1%	Score 154.5	DB 5	Length 527
-------------	-------	-------------	------	------------

Best Local Similarity 39.6%; Pred. NO. 9.8e-08;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4



TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 562 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-811-949-43

Query Match 11.1%; Score 154.5; DB 2; length 562;  
Best Local Similarity 39.6%; Pred. No. 1.1e-07;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 25 CFMDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 75  
DB 127 CYEDGISTRGTWSTAESGAECTNW--NSSALAQKPYSGRRPDAIRLGLGNHNYCRNPDR 184

QY 76 DPRGPWCYVSGEAGVPEKRPCEDLRCPEPTS 106  
DB 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

RESULT 14  
US-08-560-098A-50  
; Sequence 50, Application US/08560098A  
; Patent No. 5976841

GENERAL INFORMATION:  
APPLICANT: MNENDT, Stephan  
APPLICANT: HEINZEL-WIELAND, Regina  
APPLICANT: STEFFENS, Gerd Josef  
TITLE OF INVENTION: Proteins having Fibrinolytic and  
NUMBER OF SEQUENCES: 60  
TITLE OF INVENTION: Coagulation-inhibiting Properties  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/560,098A  
FILING DATE: 17-NOV-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: P 44 40 892.7  
FILING DATE: 17-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 148/42448  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 562 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-560-098A-50

Query Match 11.1%; Score 154.5; DB 2; length 562;  
Best Local Similarity 39.6%; Pred. No. 1.1e-07;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 25 CFMDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 75

DB 127 CYEDGISTRGTWSTAESGAECTNW--NSSALAQKPYSGRRPDAIRLGLGNHNYCRNPDR 184  
QY 76 DPRGPWCYVSGEAGVPEKRPCEDLRCPEPTS 106  
DB 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

RESULT 15  
US-08-883-795A-38  
; Sequence 38, Application US/08883795A  
; Patent No. 5985607

GENERAL INFORMATION:  
APPLICANT: Delcive, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
TITLE OF INVENTION: Vectors for Tissue Plasmidogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,795A  
FILING DATE: 27-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Micheline  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 364-7311

INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 562 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein (TPA)  
US-08-883-795A-38

Query Match 11.1%; Score 154.5; DB 2; length 562;  
Best Local Similarity 39.6%; Pred. No. 1.1e-07;  
Matches 36; Conservative 6; Mismatches 36; Indels 13; Gaps 4;

QY 25 CFMDNGHLYREDQTSAPGLRCLNWLDAQSGLASAPVS-----GAGNHSYCRNPDE 75  
DB 127 CYEDGISTRGTWSTAESGAECTNW--NSSALAQKPYSGRRPDAIRLGLGNHNYCRNPDR 184  
QY 76 DPRGPWCYVSGEAGVPEKRPCEDLRCPEPTS 106  
DB 185 DSK-PWCYVF-KAGKYSSEFCSTPACSEGNS 213

Search completed: November 25, 2003, 13:34:22  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:38:09 ; Search time 21 seconds

(without alignments)  
1204.398 Million cell updates/sec

Title: US-10-057-951-2

Perfect score: 263

Sequence: 1 MLAWVQAFVSNMLLAAY.....PVDPGSGTPTLMGQACTPGA 263

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 100 summaries

Database :

1: pir76:\*  
2: pir1:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	9	3.4	331	2	G90037
2	8	3.0	429	2	S23258
3	8	3.0	602	2	G97293
4	8	3.0	716	1	A40332
5	8	3.0	716	1	JC5061
6	7	3.0	771	2	C70716
7	8	3.0	3507	2	T34513
8	7	2.7	41	2	S01815
9	7	2.7	96	2	JC2561
10	7	2.7	120	2	S63308
11	7	2.7	144	2	S08284
12	7	2.7	144	2	S74403
13	7	2.7	150	2	AC1060
14	7	2.7	154	2	F69520
15	7	2.7	164	2	A96594
16	7	2.7	179	2	G75594
17	7	2.7	222	2	C75342
18	7	2.7	226	2	D69780
19	7	2.7	228	2	S57610
20	7	2.7	229	2	A61133
21	7	2.7	229	2	F40439
22	7	2.7	273	2	F87414
23	7	2.7	286	2	F89796
24	7	2.7	291	2	D9460
25	7	2.7	301	2	S57923
26	7	2.7	303	2	T28999
27	7	2.7	314	2	T32247
28	7	2.7	326	2	AD2612
29	7	2.7	326	2	B97394

30	7	2.7	332	2	AH1994	hypothetical prote
31	7	2.7	335	2	F64080	glycerol-3-phospha
32	7	2.7	335	2	D64397	hypothetical prote
33	7	2.7	346	2	AB2129	iron(III) dicitrat
34	7	2.7	351	1	S76613	N-acetyl-gamma-glu
35	7	2.7	356	2	A86590	hypothetical prote
36	7	2.7	356	2	D72033	hypothetical prote
37	7	2.7	365	2	F72033	hypothetical prote
38	7	2.7	365	2	B81505	hypothetical prote
39	7	2.7	365	2	C86590	hypothetical prote
40	7	2.7	388	2	S18560	hypothetical prote
41	7	2.7	398	2	B83252	xylr protein - lac
42	7	2.7	421	2	T43406	probable amino aci
43	7	2.7	448	2	AB2740	cullin-3 - fission
44	7	2.7	448	2	H97520	acetyl-CoA carboxy
45	7	2.7	448	2	T06698	biotin carboxylase
46	7	2.7	463	2	AD0799	hypothetical prote
47	7	2.7	483	2	A53918	probable membrane
48	7	2.7	485	2	T03638	chitinase (EC 3.2.
49	7	2.7	485	2	S54116	hypothetical prote
50	7	2.7	511	2	AB0397	hypothetical prote
51	7	2.7	512	2	G65048	multidrug resistan
52	7	2.7	512	2	D91072	multidrug resistan
53	7	2.7	512	2	G85916	multidrug resistan
54	7	2.7	512	2	AG0842	hypothetical prote
55	7	2.7	520	2	AL1295	multidrug resistan
56	7	2.7	528	1	WHHUY4	adenylosuccinate s
57	7	2.7	540	2	H86777	tyrosine 3-monooxy
58	7	2.7	567	1	D43719	hypothetical prote
59	7	2.7	575	1	S35138	urease (EC 3.5.1.5
60	7	2.7	585	2	S74477	acetylactate synth
61	7	2.7	591	2	S77707	hypothetical prote
62	7	2.7	592	2	S54489	phosphoribosylamin
63	7	2.7	629	2	C87048	phosphoribosylamin
64	7	2.7	631	2	T15370	probable ABC trans
65	7	2.7	701	2	C97910	hypothetical prote
66	7	2.7	710	1	I51283	ATP-dependent prote
67	7	2.7	785	2	T38359	hepatocyte growth
68	7	2.7	798	2	S62405	cullin 3 homolog -
69	7	2.7	937	2	A45082	hypothetical prote
70	7	2.7	1101	2	G70951	neurotrophic recep
71	7	2.7	1157	1	S49247	probable ATP-depen
72	7	2.7	1296	1	T16859	paracrystal
73	7	2.7	1538	2	H70846	hypothetical glyci
74	7	2.7	1638	2	T30313	hypothetical glyci
75	7	2.7	2472	2	E83594	chemotaxis protein
76	7	2.7	3433	1	GMWVKY	still frameshift p
77	6	2.3	38	2	E86077	genome polyprotein
78	6	2.3	41	2	A42064	hypothetical prote
79	6	2.3	49	2	B64323	lactam utilization
80	6	2.3	69	2	H83236	hypothetical prote
81	6	2.3	71	2	B84284	cold acclimation p
82	6	2.3	73	2	AD1043	hypothetical prote
83	6	2.3	73	2	T14887	transcription regu
84	6	2.3	74	2	D82753	hypothetical prote
85	6	2.3	78	2	AG2814	hypothetical prote
86	6	2.3	81	2	E70774	30S ribosomal prot
87	6	2.3	81	2	T09979	probable atp prot
88	6	2.3	82	2	A97820	H+-transporting tw
89	6	2.3	83	1	C42645	hypothetical prote
90	6	2.3	85	2	F81664	ribosomal protein
91	6	2.3	86	2	G70661	ribosomal protein
92	6	2.3	86	2	H72054	hypothetical prote
93	6	2.3	89	2	G86570	ribosomal protei
94	6	2.3	89	2	A60140	s17 ribosomal prot
95	6	2.3	93	2	AF1018	plasmid (EC 3.4.21
96	6	2.3	95	2	D64361	probable membrane
97	6	2.3	99	2	C70941	hypothetical prote
98	6	2.3	100	2	D87013	hypothetical prote
99	6	2.3	102	2	D90203	conserved hypothet
100	6	2.3	105	2	D83243	ATP synthase subun
						hypothetical prote

## ALIGNMENTS

## RESULT 1

hypothetical protein SA2162 [imported] - Staphylococcus aureus (strain N315)  
 C/Species: Staphylococcus aureus  
 C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C/Accession: G90037  
 R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A/Reference number: A89758; MUID:21311952; PMID:11418146  
 A/Accession: G90037  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-331 <KIR>  
 A/Cross-references: GB:BA000016; PTD:g13702323; PIDN:BAB43464.1; GSPDB:GN00149  
 A/Experimental source: Strain N315  
 C/Genetics:  
 A/Gene: SA2162

Query Match 3.4%; Score 9; DB 2; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 0.65;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 IIAIGAGIT 189  
 |||||  
 Db 97 IIAIGAGIT 105

## RESULT 2

adenylosuccinate synthase (EC 6.3.4.4) - Thiobacillus ferrooxidans  
 N/Alternate names: IMP-ascpartate ligase  
 C/Species: Thiobacillus ferrooxidans  
 C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
 C/Accession: S23258  
 R/Kusano, T.; Takeshima, T.; Sugawara, K.; Inoue, C.; Shiratori, T.; Yano, T.; Fukumori,  
 J. Biol. Chem. 267, 11242-11247, 1992  
 A>Title: Molecular cloning of the gene encoding Thiobacillus ferrooxidans Fe(II) oxidase  
 A/Reference number: S23258; MUID:92283830; PMID:1317860  
 A/Accession: S23258  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-429 <KUS>  
 A/Cross-references: EMBL:X57324; NID:g48167; PIDN:CAA0593.1; PID:g48168  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1991  
 C/Genetics:  
 A/Gene: pura  
 C/Function:  
 C/Complex: homodimer  
 A/Description: catalyzes the formation of AMP (with GDP and phosphate) from GTP, IMP, and  
 A/Pathway: AMP biosynthesis; purine nucleotide biosynthesis (the first enzyme in the AMP  
 C/Superfamily: adenylosuccinate synthase  
 C/Keywords: AMP biosynthesis; GTP binding; homodimer; ligase; purine nucleotide biosynth  
 F:14/Binding site: GMP (Lys) #status predicted

Query Match 3.0%; Score 8; DB 2; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 LGYVLGIT 176  
 |||||  
 Db 260 LGYVLGIT 267

## RESULT 3

G97293  
 ATP-dependent Zn protease, FTSH [imported] - Clostridium acetobutylicum  
 C/Species: Clostridium acetobutylicum

C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C/Accession: G97293

R/Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; I  
 .; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001

A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
 A/Reference number: A96900; MUID:21359325; PMID:21359325

A/Accession: G97293

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-602 <KIR>

A/Cross-references: GB:AE001437; PIDN:AA681138.1; PID:g15026273; GSPDB:GN00168

A/Experimental source: Clostridium acetobutylicum ATCC824

C/Genetics:  
 A/Gene: CAC3202

C/Superfamily: cell division protein ftsH, FtsH/SEC18/CDC48-type ATP-binding domain

Query Match 3.0%; Score 8; DB 2; Length 602;  
 Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 VSGEAGVP 91  
 |||||  
 Db 213 VSGEAGVP 220

## RESULT 4

macrophage-stimulating protein 1 precursor - mouse  
 N/Alternate names: hepatocyte growth factor-like protein  
 C/Species: Mus musculus (house mouse)  
 C/Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 18-Jun-1999  
 C/Accession: A40332; B40332  
 R/Degen, S.J.F.; Stuart, L.A.; Han, S.; Jamison, C.S.  
 Biochemistry 30, 9781-9791, 1991

A>Title: Characterization of the mouse cDNA and gene coding for a hepatocyte growth f  
 A/Reference number: A40332; MUID:92002017; PMID:1832957

A/Accession: A40332

A/Molecule type: DNA

A/Residues: 1-716 <DEG>

A/Cross-references: GB:M74180; NID:g193831; PIDN:AA50166.1; PID:g193832

A/Accession: B40332

A/Molecule type: mRNA

A/Residues: 1-18, 'P', 20-716 <DEG2>

A/Cross-references: GB:M74181; NID:g193833; PIDN:AA50167.1; PID:g193834

C/Genetics:  
 A/Insertions: 18/1; 67/2; 105/1; 143/2; 189/1; 229/2; 269/1; 334/2; 378/1; 412/2; 458/1

C/Complex: disulfide-bonded heterodimer of chains derived from the same precursor

C/Superfamily: hepatocyte growth factor; kirtling homology; trypsin homology

C/Keywords: duplication; glycoprotein; growth factor; kirtling

F:1-31/Domain: signal sequence #status predicted <SIG>

F:19-488,489-716/Product: macrophage-stimulating protein 1 #status experimental <MAT>

F:110-186/Domain: alpha chain #status experimental <ACH>

F:191-268/Domain: kirtling homology <KR1>

F:292-370/Domain: kirtling homology <KR2>

F:379-457/Domain: kirtling homology <KR3>

F:464-711/Domain: beta chain #status experimental <BCB>

F:469-709/Domain: trypsin homology <TRY>

F:72,173,305,620/Binding site: carbonylcarate (Asn) (covalent) #status predicted

Query Match 3.0%; Score 8; DB 1; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 DPGPCWCY 83  
 |||||  
 Db 163 DPGPCWCY 170

## RESULT 5

UC5061  
 macrophage-stimulating protein 1 precursor - rat



C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 16-Jun-2000  
 C/Accession: J05061  
 R:Oshiro, K.; Iwama, A.; Matsuno, K.; Esaki, T.; Sakamoto, O.; Hamaguchi, I.; Takasu, N.  
 Biochem. Biophys. Res. Commun. 227, 273-280, 1996  
 A>Title: Molecular cloning of rat macrophage-stimulating protein and its involvement in  
 A/Reference number: J05061; MUID:97011126; PMID:8858136  
 A/Accession: J05061  
 A:Molecule type: mRNA  
 A:Residues: 1-716 <OHS>  
 A/Cross-references: EMBL:X55096; NID:g1669718; PIDN:CAA64473.1; PID:g1669719  
 C/Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
 C/Superfamily: hepatocyte growth factor; kringie homology; trypsin homology  
 C/Keywords: duplication; glycoprotein; growth factor; kringie  
 F:1-31/Domain: signal sequence #status predicted <SIG>  
 F:32-48/Domain: Product: macrophage-stimulating protein 1 #status predicted <MAT>  
 F:52-48/Domain: macrophage-stimulating protein 1 alpha chain #status predicted <ACH>  
 F:110-186/Domain: kringie homology <KR11>  
 F:191-268/Domain: kringie homology <KR12>  
 F:292-370/Domain: kringie homology <KR13>  
 F:379-457/Domain: kringie homology <KR14>  
 F:489-716/Domain: macrophage-stimulating protein 1 beta chain #status predicted <BGH>  
 F:489-709/Domain: trypsin homology <TRY>  
 F:72,305,620/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 3.0%; Score 8; DB 1; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 76 DRGPWCY 83  
 Db 163 DRGPWCY 170

RESULT 6  
 C70716  
 probable DNA helicase - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C/Accession: C70716

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.  
 ; Raftery, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; MUID:98295987; PMID:9634230  
 A/Accession: C70716  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-771 <COL>  
 A/Cross-references: GB:I29700; GB:AL123456; NID:g3261628; PIDN:CAB02001.1; PID:g1524213  
 A/Experimental source: strain H37Rv  
 C/Genetics:

A:Gene: uvrd  
 C:Superfamily: helicase II

Query Match 3.0%; Score 8; DB 2; Length 771;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 SAPVSGAG 65  
 Db 692 SAPVSGAG 699

RESULT 7  
 T34513  
 hypothetical protein ZK783.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C/Accession: T34513  
 R:Favella, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994  
 A/Description: The sequence of C. elegans cosmid ZK783.  
 A/Reference number: Z21536  
 A/Accession: T34513

A>Status: preliminary; translated from GB/EMBL/DDBT  
 A:Molecule type: DNA  
 A:Residues: 1-3507 <FAV>  
 A/Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1  
 A/Experimental source: strain Bristol N2; clone ZK783  
 C/Genetics:  
 A:Gene: CESP:ZK783.1  
 A:Map position: 3  
 A:Insertions: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match 3.0%; Score 8; DB 2; Length 3507;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 85 SGEAGVPE 92  
 Db 1269 SGEAGVPE 1276

RESULT 8  
 S01815  
 hemoglobin BII1 - tube worm (Lamellibrachia sp.) (fragment)  
 C:Species: Lamellibrachia sp.  
 C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 04-Mar-2000  
 C/Accession: S01815  
 R:Suzuki, T.; Takagi, T.; Ohta, S.  
 Biochem. J. 255, 541-545, 1988  
 A>Title: N-terminal amino acid sequence of the deep-sea tube worm haemoglobin remarks  
 A/Reference number: S01807; MUID:89076216; PMID:3202832  
 A/Accession: S01815  
 A:Molecule type: protein  
 A:Residues: 1-41 <SUZ>  
 C/Superfamily: globin; globin homology  
 C/Keywords: oxygen carrier

Query Match 2.7%; Score 7; DB 2; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AEAYSG 23  
 Db 16 AEAYSG 22

RESULT 9  
 JC2561  
 chaperonin groESX protein - Amoeba proteus  
 C:Species: Amoeba proteus  
 C>Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 26-Aug-1999  
 C/Accession: JC2561

R:Ahm, T.I.; Lim, S.T.; Leeu, H.K.; Lee, J.E.; Jeon, K.W.  
 Gene 148[128], 43-49, 1994  
 A>Title: A novel strong promoter of the groEX operon of symbiotic bacteria in Amoeba

A/Note: due to a typographical error the volume number 148 appears as 128  
 A/Reference number: JC2561  
 A/Accession: JC2561  
 A:Molecule type: DNA  
 A:Residues: 1-96 <AHN>  
 A/Cross-references: GB:M6549; NID:g155400; PIDN:AAC09380.1; PID:g155401

C/Comment: This protein is involved in the assembly of oligomeric protein complexes,  
 C/Genetics:  
 A:Gene: groESX  
 C/Superfamily: chaperonin groES  
 C/Keywords: molecular chaperone

Query Match 2.7%; Score 7; DB 2; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 IIAIGAG 187  
 DB 40 IIAIGAG 46

RESULT 10  
 S69308  
 probable membrane protein YLR302c - yeast (*Saccharomyces cerevisiae*)  
 N:Alternate names: hypothetical protein L8003.2-b  
 C:Species: *Saccharomyces cerevisiae*  
 C:Date: 20-Jul-1996 #sequence\_revision 23-Aug-1996 #text\_change 19-Apr-2002  
 C:Accession: S69308  
 R:Pauley, A.  
 submitted to the EMBL Data Library, November 1994  
 A:Description: The sequence of *S. cerevisiae* cosmid 8003.  
 A:Reference number: S50366  
 A:Accession: S69308  
 A:Molecule type: DNA  
 A:Residues: 1-120 <PAU>  
 A:Cross-references: EMBL:U17243; NID:G596030; PID:G2340968; GSPDB:GN00012; MIPS:YLR302C  
 C:Genetics:  
 A:Gene: MIPS:YLR302C  
 A:Cross-references: SGD:S0004293  
 A:Map position: 12R  
 C:Superfamily: *Saccharomyces cerevisiae* probable membrane protein YLR302C  
 C:Keywords: transmembrane protein  
 F:44-60/Domain: transmembrane #status predicted <TMM>

Query Match 2.7%; Score 7; DB 2; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EIVDEKT 235  
 DB 25 EIVDEKT 31

RESULT 11  
 S08284  
 hemoglobin A111 - tube worm (*Lamellibrachia* sp.)  
 C:Species: *Lamellibrachia* sp.  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 04-Mar-2000  
 C:Accession: S08284; S01809  
 R:Suzuki, T.; Takagi, T.; Ohka, S.  
 Biochem. J. 266, 221-225, 1990  
 A:Title: Primary structure of a constituent polypeptide chain (A111) of the giant haemoglobin  
 A:Reference number: S08284; MUID:90179711; PMID:2310374  
 C:Accession: S08284  
 A:Molecule type: protein  
 A:Residues: 1-144 <SUZ>  
 R:Suzuki, T.; Takagi, T.; Ohka, S.  
 Biochem. J. 255, 541-545, 1988  
 A:Title: N-terminal amino acid sequence of the deep-sea tube worm haemoglobin remarkably  
 A:Reference number: S01807; MUID:89076216; PMID:3202832  
 C:Accession: S01809  
 A:Molecule type: protein  
 A:Residues: 1-24 <SUZ>  
 C:Superfamily: globin; globin homology  
 C:Keywords: chromoprotein; heme; iron; metalloprotein; oxygen carrier  
 F:95/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 2.7%; Score 7; DB 2; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AEAYGSG 23  
 DB 16 AEAYGSG 22

RESULT 12  
 S74403

hypothetical protein slr0491 - *Synechocystis* sp. (strain PCC 6803)  
 C:Species: *Synechocystis* sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C:Accession: S74403  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yae  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
 s.  
 A:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S74403  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-144 <KAN>  
 A:Cross-references: EMBL:D64001; GB:AB001339; NID:G1001102; PIDN:BA010321.1; PID:d10  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 2.7%; Score 7; DB 2; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TLGYVLG 174  
 DB 40 TLGYVLG 46

RESULT 13  
 AC1060  
 conserved hypothetical protein STY4806 [imported] - *Salmonella enterica* subsp. enter  
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
 A>Note: this species has also been called *Salmonella typhi*  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AC1060  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church  
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* s  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 C:Accession: AC1060  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-150 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD06928.1; PID:G16505576; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY4806  
 C:Superfamily: hypothetical protein H10227

Query Match 2.7%; Score 7; DB 2; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 VDERTVV 237  
 DB 107 VDERTVV 113

RESULT 14  
 F69520  
 conserved hypothetical protein AF216 - *Archaeoglobus fulgidus*  
 C:Species: *Archaeoglobus fulgidus*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: F69520  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Do  
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kitzness,  
 Glodek, A.; Zhou, L.; Overbeek, R.; Dacey, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Uitterlind, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
 A:Reference number: A69250; MUID:96049343; PMID:9389475

A:Accession: F69520  
 A:Structure: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-154 <KLE>  
 A:Cross-references: GB:AE000955; GB:AE000782; NID:G2689278; PIDN:AAB89089.1; PID:G264836

Query Match 2.7%; Score 7; DB 2; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 GILIGYS 193  
 |||||  
 DB 104 GILIGYS 110

## RESULT 15

A96594  
 hypothetical protein F7A10.6 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: A96594  
 R:Neologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.; Hansen, N.F.; Hughes, B.; Huizcar, L.  
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Sato, Y.; Liu, X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani, A.; Authors: Salberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, K.; M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: A96594

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-164 <STD>

A:Cross-references: GB:AE005173; NID:G10645434; PIDN:AAG21551.1; GSPDB:GND0141

C:Genetics:

A:Gene: F7A10.6

A:Map position: 1

Query Match 2.7%; Score 7; DB 2; Length 164;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 KKDGLT 169  
 |||||  
 DB 145 KKDGLT 151

## RESULT 16

G75594

cobinamide kinase/cobinamide phosphate guanylyltransferase - Deinococcus radiodurans (str

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C:Accession: G75594

R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: G75594

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-179 <WHI>

A:Cross-references: GB:AE001862; GB:AE001825; NID:G6460466; PIDN:AAF12282.1; PID:G646057

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRA0020

C:Superfamily: cobinamide kinase

Query Match 2.7%; Score 7; DB 2; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VSNMLLA 17  
 |||||  
 DB 96 VSNMLLA 102

## RESULT 17

C75342

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C:Accession: C75342

R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: C75342

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-222 <WHI>

A:Cross-references: GB:AE002027; GB:AE000513; NID:G6459655; PIDN:AAF11429.1; PID:G645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1975

A:Map position: 1

Query Match 2.7%; Score 7; DB 2; Length 222;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 LASAPVS 62  
 |||||  
 DB 2 LASAPVS 8

## RESULT 18

D69780

hypothetical protein ydf - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C:Accession: D69780

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Ertian, K.D.; Ertington, J.; Fabre, C.; Ferrati,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koehler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lario

A:Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogikawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scari

A:Authors: Schlicht, S.; Schroefer, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se

keuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpiltra, F.; Tognoni, A.; Tosato, V.; Uchida

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A69580; MUID:98044033; PMID:9383377

A:Accession: D69780

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-226 <KUN>

A:Cross-references: GB:Z99106; GB:AL009126; NID:G2632653; PIDN:CAH12346.1; PID:el1825

A:Experimental source: strain 168

C:Genetics:

A:Gene: ydf

Query Match 2.7%; Score 7; DB 2; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 SKEKKDL 166  
 |||||  
 Db 99 SKEKKDL 105

RESULT 19  
 granula associated protein 24 - Alcaligenes eutrophus  
 C/Species: Alcaligenes eutrophus  
 C/Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 09-Sep-1997  
 C/Accession: S57610  
 R/Wong, R.; Pries, A.; Steinhuechel, A.; Mayer, F.  
 submitted to the EMBL Data Library, March 1995  
 A/Description: Analysis of a 24 kDa protein associated with the polyhydroxyalkanoic acid  
 A/Reference number: S57610  
 A/Accession: S57610  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-228 <MT>  
 A/Cross-references: EMBL:X85729; NID:g886423; PID:g886424

Query Match 2.7%; Score 7; DB 2; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 AAAYQPV 149  
 |||||  
 Db 69 AAAYQPV 75

RESULT 20  
 prolactin precursor - turkey  
 C/Species: Melagris gallioavo (common turkey)  
 C/Date: 10-Mar-1994 #sequence\_revision 07-Apr-1994 #text\_change 11-May-2000  
 C/Accession: A61133; S10170; A61528  
 R/Wong, E.A.; Ferrin, N.H.; Silsby, J.L.; El Halawani, M.E.  
 Gen. Comp. Endocrinol. 83, 18-26, 1991  
 A/Title: Cloning of a turkey prolactin cDNA: expression of prolactin mRNA throughout the  
 A/Reference number: A61133; MUID:91348480; PMID:1879669  
 A/Accession: A61133  
 A/Molecule type: mRNA  
 A/Residues: 1-155, R, 157-229 <MO2>  
 A/Cross-references: GB:U05952; NID:g454094; PIDN:AB60604.1; PID:g454095  
 R/Karatzas, C.N.; Zadworny, D.; Kuhnlein, U.  
 Nucleic Acids Res. 18, 3071, 1990  
 A/Title: Nucleotide sequence of turkey prolactin  
 A/Reference number: S10170; MUID:90272435; PMID:2349117  
 A/Accession: S10170  
 A/Molecule type: mRNA  
 A/Residues: 21-229 <KAR>  
 A/Cross-references: EMBL:X51769; NID:g64095; PIDN:CAA36071.1; PID:g64096  
 R/Cocoran, D.H.; Proudman, J.A.  
 Comp. Biochem. Physiol. B 99, 563-570, 1991  
 A/Title: Isoforms of turkey prolactin: evidence for differences in glycosylation and in  
 A/Reference number: A61528; MUID:92119931; PMID:1769204  
 A/Accession: A61528  
 A/Molecule type: protein  
 A/Residues: 31-70 <COR>  
 C/Superfamily: prolactin  
 C/Keywords: hormone; pituitary  
 F,1-30/Domain: signal sequence #status predicted <SIG>  
 F,31-229/Product: prolactin #status predicted <MT>  
 F,34-41,88-204,221-229/Disulfide bonds: #status predicted

Query Match 2.7%; Score 7; DB 2; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LVSNMIL 16  
 |||||

Db 17 LVSNMIL 23

RESULT 21  
 T40439  
 hypothetical protein SPBC409.12c - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C/Accession: T40439  
 R/Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Chillingworth, T.; Churcher, C.M.  
 submitted to the EMBL Data Library, August 1999  
 A/Reference number: Z21929  
 A/Accession: T40439  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-229 <LYN>  
 A/Cross-references: EMBL:AL109822; PIDN:CA852614.1; GSPDB:GN00067; SPDB:SPBC409.12c  
 C/Genetics:  
 A/Gene: SPDB:SPBC409.12c  
 A/Map position: 2

Query Match 2.7%; Score 7; DB 2; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 ITLPLSA 222  
 |||||  
 Db 192 ITLPLSA 198

RESULT 22  
 F87414  
 phage SP01 DNA polymerase-related protein [imported] - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: F87414  
 R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: F87414  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-273 <SNO>  
 A/Cross-references: GB:AE005673; NID:913422678; PIDN:AAK23314.1; GSPDB:GN00148  
 C/Genetics:  
 A/Gene: CCI333

Query Match 2.7%; Score 7; DB 2; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 EGPQADE 127  
 |||||  
 Db 123 EGPQADE 129

RESULT 23  
 F89796  
 hypothetical protein SA0305 [imported] - Staphylococcus aureus (strain N315)  
 C/Species: Staphylococcus aureus  
 C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C/Accession: F89796  
 R/Kuroda, M.; Ohba, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.;  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, I.  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A/Reference number: A89758; MUID:21311952; PMID:11181846  
 A/Accession: F89796

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-286 <KUR>  
 A:Cross-references: GB:BA000018; PTD:gl3700231; PIDN:BA841529.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA0305

Query Match 2.7%; Score 7; DB 2; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 QPVIGIS 153  
 |||||  
 Db 59 QPVIGIS 65

RESULT 24  
 D90460  
 conserved hypothetical protein [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C:Accession: D90460  
 R:Shen, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-  
 Jiong, T.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.  
 arrett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.  
 Submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: D90460  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-291 <KUR>  
 A:Cross-references: GB:AB06641; NID:g13816184; PIDN:AAK42939.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: SS02829

Query Match 2.7%; Score 7; DB 2; Length 291;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 VIIAIG 185  
 |||||  
 Db 50 VIIAIG 56

RESULT 25  
 S57923  
 SEC14 protein - yeast (Candida albicans)  
 C:Species: Candida albicans  
 C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 26-Feb-1998  
 C:Accession: S57923; S72193  
 R:Monteoliya, L.; Sanchez, M.; Pla, J.; Gil, C.; Nombela, C.  
 Submitted to the EMBL data library, September 1994  
 A:Description: Characterisation of the Candida albicans SEC14 homolog gene.  
 A:Reference number: S57923  
 A:Accession: S57923  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-301 <MON>  
 A:Cross-references: EMBL:X81937  
 R:Monteoliya, L.; Sanchez, M.; Pla, J.; Gil, C.; Nombela, C.  
 Yeast 12, 1097-1105, 1996  
 A:Title: Cloning of Candida albicans SEC14 gene homologue coding for a putative essential  
 A:Reference number: S72193; MUID:97051600; PMID:8896277  
 A:Accession: S72193  
 A:Molecule type: DNA  
 A:Residues: 1-301 <MON>  
 A:Cross-references: EMBL:X81937  
 A:Note: the authors translated the codon CTG for residue 180 as Ser  
 C:Genetics:  
 A:Gene: SEC14  
 C:Superfamily: cellular retinaldehyde-binding protein; cellular retinaldehyde-binding pr

F:59-266/Domain: cellular retinaldehyde-binding protein homology <CRB>  
 Query Match 2.7%; Score 7; DB 2; Length 301;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ILGYSYK 195  
 |||||  
 Db 242 ILGYSYK 248

RESULT 26  
 T28999  
 hypothetical protein ZC513.8 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T28999  
 R:Wu, X.; Le, T.T.  
 submitted to the EMBL Data Library, April 1996  
 A:Description: The sequence of C. elegans cosmid ZC513.  
 A:Reference number: Z20551  
 A:Accession: T28999  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-303 <MON>  
 A:Cross-references: EMBL:U53155; PIDN:AA048270.1; GSPDB:GN00023; CESP:ZC513.8  
 A:Experimental source: strain Bristol N2; clone ZC513  
 C:Genetics:  
 A:Gene: CESP:ZC513.8  
 A:Map position: 5  
 A:Introns: 55/3; 229/1  
 C:Superfamily: unassigned collagens

Query Match 2.7%; Score 7; DB 2; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 GOAGTPG 262  
 |||||  
 Db 274 GOAGTPG 280

RESULT 27  
 T32247  
 hypothetical protein T15B7.5 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T32247  
 R:Pauley, A.; Gattung, S.  
 submitted to the EMBL Data Library, September 1997  
 A:Description: The sequence of C. elegans cosmid T15B7.  
 A:Reference number: Z21139  
 A:Accession: T32247  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-314 <PAU>  
 A:Cross-references: EMBL:AF022985; PIDN:AB69960.1; GSPDB:GN00023; CESP:T15B7.5  
 A:Experimental source: strain Bristol N2; clone T15B7  
 C:Genetics:  
 A:Gene: CESP:T15B7.5  
 A:Map position: 5  
 A:Introns: 273/1  
 C:Superfamily: unassigned collagens

Query Match 2.7%; Score 7; DB 2; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 GOAGTPG 262  
 |||||  
 Db 251 GOAGTPG 257

RESULT 28  
AD2612  
Iron-sulfur cluster binding protein [imported] - *Agrobacterium tumefaciens* (strain C58,  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AD2612  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; Mclell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AD2612  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-326 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AA141314.1; PID:g17738625; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu0292  
A:Map position: circular chromosome

Query Match 2.7%; Score 7; DB 2; Length 326;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 PDEDPRG 79  
Db 33 PDEDPRG 39

RESULT 29  
B97394  
Hypothetical protein AGR\_C\_502 [imported] - *Agrobacterium tumefaciens* (strain C58, Cere  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C:Accession: B97394  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Mollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: B97394  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-326 <KUR>  
A:Cross-references: GB:AE007863; PIDN:AKK6107.1; PID:g15155190; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_502  
A:Map position: circular chromosome

Query Match 2.7%; Score 7; DB 2; Length 326;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 PDEDPRG 79  
Db 33 PDEDPRG 39

RESULT 30  
AH1994  
Hypothetical protein all1509 [imported] - *Nostoc* sp. (strain PCC 7120)  
C:Species: *Nostoc* sp. PCC 7120  
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AH1994  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,  
Nakazawa, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing *Cyanobacterium* And

A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH1994  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-332 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA87875.1; PID:g17135330; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all1509

Query Match 2.7%; Score 7; DB 2; Length 332;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 IATGAGI 188  
Db 221 IATGAGI 227

RESULT 31  
F64080  
Glycerol-3-phosphate dehydrogenase (NAD) (EC 1.1.1.8) - *Haemophilus influenzae* (stra  
C:Species: *Haemophilus influenzae*  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002  
C:Accession: F64080  
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavages  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
; D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Ventec  
A:Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: F64080  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-335 <TTGR>  
A:Cross-references: GB:U32743; GB:L42023; NID:g1573597; PIDN:ACC22264.1; PID:g1573598  
A:Superfamily: glycerol-3-phosphate dehydrogenase  
C:Keywords: oxidoreductase

Query Match 2.7%; Score 7; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 IATGAGI 188  
Db 198 IATGAGI 204

RESULT 32  
D64397  
Hypothetical protein MJ0780 - *Methanococcus jannaschii*  
C:Species: *Methanococcus jannaschii*  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: D64397  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Flieschmann, R.D.; Sutton, G.G.; Blak  
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Woese  
A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannasc*  
A:Reference number: A64300; MUID:96317999; PMID:8688087  
A:Accession: D64397  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-335 <BUL>  
A:Cross-references: GB:U67522; GB:L77117; NID:g2826315; PIDN:AA898779.1; PID:g14996000  
C:Genetics:  
A:Map position: REV01539-700532

Query Match 2.7%; Score 7; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 IGAGIIL 190  
 |||||  
 Db 137 IGAGIIL 143

## RESULT 33

AB2129  
 Iron(III) dicitrate transport system permease protein al12585 [imported] - Nostoc sp. (S  
 C:Species: Nostoc sp. strain PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AB2129  
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8: 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AB2129  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-346 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA074284.1; PID:q17131678; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all12585  
 C:Superfamily: vitamin B12 transport protein btuc

Query Match 2.7%; Score 7; DB 2; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 217 TLPLSAR 223  
 |||||  
 Db 133 TLPLSAR 139

## RESULT 34

S76613  
 N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) - Synecchocystis sp. (strain PC  
 C:Species: Synecchocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: S76613  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3: 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis  
 S.  
 A:Reference number: S74322; MUID:37061201; PMID:8905231  
 A:Accession: S76613  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-351 <KAN>  
 A:Cross-references: EMBL:D64004; GB:AB001339; NID:q1001701; PIDN:BA10557.1; PID:q100172  
 C:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Superfamily: N-acetyl-gamma-glutamyl-phosphate reductase  
 C:Keywords: oxidoreductase

Query Match 2.7%; Score 7; DB 1; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 13 NMLIAEA 19  
 |||||  
 Db 195 NMLIAEA 201

## RESULT 35

A86590  
 hypothetical protein CPJ0795 [imported] - Chlamydia pneumoniae (strain J138)  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C:Accession: A86590  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
 Nucleic Acids Res. 28: 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
 A:Reference number: A86491; MUID:20330349; PMID:10871362  
 A:Accession: A86590  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-356 <SPO>  
 A:Cross-references: GB:BA000008; NID:98979168; PIDN:BA99003.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: CPJ0795

Query Match 2.7%; Score 7; DB 2; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 164 KDGLTIG 170  
 |||||  
 Db 2 KDGLTIG 8

## RESULT 36

D72033  
 hypothetical protein CP1076 [imported] - Chlamydia pneumoniae (strains CWL029 and  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 C:Accession: D72033; D81505  
 R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,  
 Nature Genet. 21: 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; MUID:99206606; PMID:10192388  
 A:Accession: D72033  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-356 <ARN>  
 A:Cross-references: GB:AE001661; GB:AE001363; NID:q4377104; PIDN:AD18933.1; PID:q4377  
 A:Experimental source: strain CWL029  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe  
 Nucleic Acids Res. 28: 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
 A:Reference number: AB1500; MUID:20150255; PMID:10684935  
 A:Accession: D81505  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-356 <RBA>  
 A:Cross-references: GB:AE002264; GB:AE002161; NID:q7169984; PIDN:AA38848.1; PID:q718  
 A:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: CPN0795; CP1076

Query Match 2.7%; Score 7; DB 2; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 164 KDGLTIG 170  
 |||||  
 Db 2 KDGLTIG 8

## RESULT 37

F72033  
 hypothetical protein - Chlamydia pneumoniae (strain CWL029)  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
 C:Accession: F72033  
 R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,  
 Nature Genet. 21: 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; MUID:99206606; PMID:10192388  
 A:Accession: F72033

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-365 <ARN>  
A;Cross-references: GB:AE001661; GB:AE001363; NID:g4377104; PIDN:AAD18935.1; PID:g4377104  
A;Experimental source: strain CWL029  
C;Genetics:  
A;Gene: CP0797

Query Match 2.7%; Score 7; DB 2; Length 365;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDUGTIG 170  
|||||  
Db 142 KDUGTIG 148

RESULT 38  
B81505  
hypothetical protein CP1074 [imported] - Chlamydomophila pneumoniae (strain AR39)  
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C;Accession: B81505  
R;Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwyn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A;Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.  
A;Reference number: AB1500; MUID:20150255; PMID:10684935  
A;Accession: B81505  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-365 <REA>  
A;Cross-references: GB:AE002264; GB:AE002161; NID:g7189984; PIDN:AAF38846.1; PID:g7189984  
A;Experimental source: strain AR39, HL cells  
C;Genetics:  
A;Gene: CP1074

Query Match 2.7%; Score 7; DB 2; Length 365;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDUGTIG 170  
|||||  
Db 142 KDUGTIG 148

RESULT 39  
C86590  
hypothetical protein CP10797 [imported] - Chlamydomophila pneumoniae (strain J138)  
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C;Accession: C86590  
R;Shirol, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise  
Nucleic Acids Res. 28, 2311-2314, 2000  
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A;Reference number: AB6491; MUID:20330349; PMID:10871362  
A;Accession: C86590  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-365 <STO>  
A;Cross-references: GB:BA000008; NID:g8979171; PIDN:BAA99005.1; GSPDB:GN00142  
A;Experimental source: strain J138  
C;Genetics:  
A;Gene: CP10797

Query Match 2.7%; Score 7; DB 2; Length 365;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDUGTIG 170  
|||||  
Db 142 KDUGTIG 148

RESULT 40  
S18560  
xylR protein - Lactobacillus pentosus

C;Species: Lactobacillus pentosus  
C;Date: 22-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 15-Oct-1999  
C;Accession: S18560  
R;Lekman, B.C.; van Santen, P.; Verdoes, J.C.; Kruse, J.; Leer, R.J.; Posno, M.; Pou  
Mol. Gen. Genet. 230, 161-169, 1991  
A;Title: Organization and characterization of three genes involved in D-xyllose catabo  
A;Reference number: S18560; MUID:92079891; PMID:1660563  
A;Accession: S18560

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-388 <LOK>  
A;Cross-references: EMBL:M57384; NID:g149604; PIDN:AAA25257.1; PID:g149605  
A;Note: the authors translated the initiation codon GTG for residue 1 as Val  
C;Genetics:  
A;Gene: xylR  
A;Start codon: GTG  
A;Superfamily: xyllose repressor; glucose kinase homology  
C;Keywords: DNA binding; transcription regulation  
P;144-267/Domain: glucose kinase homology <GKH>

Query Match 2.7%; Score 7; DB 2; Length 388;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 IGAGITL 190  
|||||  
Db 219 IGAGITL 225

RESULT 41  
B83252  
probable amino acid aminotransferase PA3139 [imported] - Pseudomonas aeruginosa (stra  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C;Accession: B83252  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Ryan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; I  
..; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A;Reference number: AB2950; MUID:20437337; PMID:10984043  
A;Accession: B83252

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-398 <STO>  
A;Cross-references: GB:AE004738; GB:AE004091; NID:g9949252; PIDN:AMG06527.1; GSPDB:GN  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA3139  
C;Superfamily: aspartate aminotransferase

Query Match 2.7%; Score 7; DB 2; Length 398;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 NALPARS 141  
|||||  
Db 169 NALPARS 175

RESULT 42  
T43406  
cullin-3 - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C;Accession: T43406  
R;Komitani, K.; Toda, T.  
submitted to the EMBL Data Library, August 1998  
A;Description: Pcu3 (S. pombe cullin-3).



A:Reference number: Z22490  
 A:Accession: T43406  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-421 <KOR>  
 A:Cross-references: EMBL:AB017026; PIDD:BAA32519.1  
 C:Genetics:  
 A:Gene: PCU3  
 C:Function:  
 A:Description: involved in the distinct stress-response pathway  
 A>Note: not a component of SCFPop1,2 like cullin-1

Query Match 2.7%; Score 7; DB 2; Length 421;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 ETTSQL 109  
 DB 218 ETTSQL 224

## RESULT 43

AB2740  
 acetyl-CoA carboxylase, biotin carboxylase [imported] - Agrobacterium tumefaciens (sera)  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C:Accession: AB2740  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erase, G.; Giller, W.; Grant, C.; Guenther, D.; Kuyavain, T.; Levy, R.; Li, M.; McClellan,  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, B.W.  
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:21608550; PMID:11743193  
 A:Accession: AB2740  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-448 <KOR>  
 A:Cross-references: GB:AE008668; PIDD:AAL42336.1; PIDD:G17739740; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: accC  
 C:Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match 2.7%; Score 7; DB 2; Length 448;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 ARSEAAA 145  
 DB 183 ARSEAAA 189

## RESULT 44

H97520  
 biotin carboxylase (a chain of acetyl-CoA carboxylase (acc) [imported] - Agrobacterium t  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C:Accession: H97520  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Onuolilo, B.; Goldm  
 A: Liu, P.; Woljam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A:Reference number: A97359; PMID:21608551; PMID:11743194  
 A:Accession: H97520  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-448 <KOR>  
 A:Cross-references: GB:AE007869; PIDD:AAK87121.1; PIDD:G15156385; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_2451

A:Map position: circular chromosome  
 C:Superfamily: biotin carboxylase; biotin carboxylase homology

QY 139 ARSEAAA 145  
 DB 183 ARSEAAA 189

## RESULT 45

T06698  
 hypotheical protein T29H11.40 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999  
 C:Accession: T06698  
 R:Quetier, F.; Choine, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Art  
 submitted to the Protein Sequence Database, April 1999  
 A:Reference number: Z15793  
 A:Accession: T06698  
 A:Molecule type: DNA  
 A:Residues: 1-448 <QUE>  
 A:Cross-references: EMBL:AL049659; GSPDB:GN00061; ATSP:T29H11.40  
 A:Experimental source: cultivar Columbia; BAC clone T29H11  
 C:Genetics:  
 A:Gene: ATSP:T29H11.40  
 A:Map position: 3  
 A:Introns: 142/3; 165/3; 205/3; 301/3; 393/3

Query Match 2.7%; Score 7; DB 2; Length 448;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SGLASAP 60  
 DB 188 SGLASAP 194

## RESULT 46

AD0799  
 probable membrane protein SRY2572 [imported] - Salmonella enterica subsp. enterica ser  
 C:Species: Salmonella enterica subsp. enterica serovar Typh  
 A>Note: this species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AD0799  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
 th, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farre  
 ; S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, J  
 A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser  
 A:Reference number: AB0502; PMID:21534947; PMID:11677608  
 A:Accession: AD0799  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-463 <PAR>  
 A:Cross-references: GB:AL513382; PIDD:CAD07574.1; PIDD:G16503566; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: SRY2572

Query Match 2.7%; Score 7; DB 2; Length 463;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 GTLGVL 173  
 DB 188 GTLGVL 194

## RESULT 47

AS3918

chitinase (EC 3.2.1.14) precursor - braconid wasp (Chelonus sp.)

C:Species: Chelonus sp.

C:Date: 28-Jul-1995 #sequence\_revision 28-Jul-1995 #text\_change 21-Jul-2000

C:Accession: A53918

R:Krishnan, A.; Nair, P.N.; Jones, D.

J. Biol. Chem. 269, 20971-20976, 1994

A:Title: Isolation, cloning, and characterization of new chitinase stored in active form

A:Reference number: A53918; MUID:94342256; PMID:8063715

A:Accession: A53918

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-483 <KRI>

A:Cross-references: GB:U00422; NID:G533504; PIDN:AAA61639.1; PID:G533505

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 2.7%; Score 7; DB 2; Length 483;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 APVSGAG 65

Db 281 APVSGAG 287

RESULT 48

T03638

hypothetical protein - maize

C:Species: Zea mays (maize)

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 20-Jun-2000

C:Accession: T03638

R:Daniel, T.J.

submitted to the EMBL Data Library, January 1995

A:Reference number: Z14985

A:Accession: T03638

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-485 <DAN>

A:Cross-references: EMBL:Z47554; NID:G1103747; PIDN:CAA87634.1

A:Experimental source: strain Black Mexican Sweet

Query Match 2.7%; Score 7; DB 2; Length 485;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 EAYGSGG 24

Db 385 EAYGSGG 391

RESULT 49

S54116

hypothetical protein - maize

C:Species: Zea mays (maize)

C:Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Sep-1997

C:Accession: S54116

R:Daniel, T.J.; Edwards, R.

submitted to the EMBL Data Library, January 1995

A:Description: Complementation of a heat shock sensitive mutant of *Escherichia coli* defi

A:Reference number: S54116

A:Accession: S54116

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-485 <DAN>

A:Cross-references: EMBL:Z47554; NID:G1103747; PID:G1360703

Query Match 2.7%; Score 7; DB 2; Length 485;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 EAYGSGG 24

Db 385 EAYGSGG 391

RESULT 50

AB0397

multidrug resistance protein B [imported] - *Yersinia pestis* (strain CO92)

C:Species: *Yersinia pestis*

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 17-May-2002

C:Accession: AB0397

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltall, R.W.; Holden, M.T.G.; Prentice, M.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel

Nature 413, 523-527, 2001

A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AB0397

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-511 <KIR>

A:Cross-references: GB:AL590842; PIDN:CAC92502.1; PID:G15981202; GSPDB:GN00175

C:Genetics:

A:Gene: emrB

C:Superfamily: lincosycin-resistance protein lmrB

Query Match 2.7%; Score 7; DB 2; Length 511;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GLASAPV 61

Db 306 GLASAPV 312

RESULT 51

G65048

multidrug resistance protein B - *Escherichia coli* (strain K-12)

C:Species: *Escherichia coli*

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: G65048; JCI345; S27558

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G65048

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-512 <BLAT>

A:Cross-references: GB:AE000353; GB:U00096; NID:G1789037; PIDN:ACT5733.1; PID:G17890

A:Experimental source: strain K-12, substrain MG1655

R:Lomovskaya, O.; Lewis, K.

Proc. Natl. Acad. Sci. U.S.A. 89, 8938-8942, 1992

A:Title: emr, an *Escherichia coli* locus for multidrug resistance.

A:Reference number: JCI344; MUID:93028382; PMID:1409590

A:Accession: JCI345

A:Molecule type: DNA

A:Residues: 1-324, 'A', 326-500, 'A', 502-512 <LOW>

A:Cross-references: GB:M86657; NID:G145834; PIDN:AAA23725.1; PID:G145836

C:Comment: This protein is resistant to carbonylcyanide m-chlorophenylhydrazone, nalidix

C:Genetics:

A:Gene: emrB

A:Map position: 57.5 min

C:Superfamily: lincosycin-resistance protein lmrB

Query Match 2.7%; Score 7; DB 2; Length 512;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GLASAPV 61

Db 307 GLASAPV 313

RESULT 52

D91072  
 multidrug resistance membrane translocase [imported] - Escherichia coli (strain O157:H7,  
 C:Species: Escherichia coli  
 C>Date: 16-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-May-2002  
 C:Accession: D91072  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gaaswara, N.; Yasunaga, T.; Kubara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
 A:Reference number: A9629; PMID:1156231; PMID:11258796  
 A:Accession: D91072  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-512 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA036971.1; PID:G13363019; GSPDB:GN00154  
 C:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: EC8548  
 C:Superfamily: lincomycin-resistance protein lmrB

Query Match 2.7%; Score 7; DB 2; Length 512;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GLASAPV 61  
 |||||  
 Db 307 GLASAPV 313

RESULT 53  
 G85916  
 hypothetical protein emrB [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
 C:Accession: G85916  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; PMID:11074935; PMID:11206551  
 A:Accession: G85916  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-512 <STO>  
 A:Cross-references: GB:AE005174; NID:G12517127; PIDN:AA057795.1; GSPDB:GN00145; UMGPR:239  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: emrB  
 C:Superfamily: lincomycin-resistance protein lmrB

Query Match 2.7%; Score 7; DB 2; Length 512;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GLASAPV 61  
 |||||  
 Db 307 GLASAPV 313

RESULT 54  
 AG0842  
 multidrug resistance protein B [imported] - Salmonella enterica subsp. enterica serovar  
 C:Species: Salmonella enterica subsp. enterica serovar typh  
 A:Note: this species has also been called Salmonella typh  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AG0842  
 R:Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Mule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AG0842  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-512 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD05926.1; PID:G16503897; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY2941  
 C:Superfamily: lincomycin-resistance protein lmrB

Query Match 2.7%; Score 7; DB 2; Length 512;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GLASAPV 61  
 |||||  
 Db 307 GLASAPV 313

RESULT 55  
 A13295  
 adenylosuccinate synthase (EC 6.3.4.4) [imported] - Brucella melitensis (strain 16M)  
 C:Species: Brucella melitensis  
 C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 17-May-2002  
 C:Accession: A13295  
 R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.U.; Patra, G.; Mijer, C.; Los, T.; Ivanov  
 Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Lec  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
 A:Reference number: AD3252; PMID:11756688  
 A:Accession: A13295  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-520 <KUR>  
 A:Cross-references: GB:AE008917; PIDN:AAU51532.1; PID:G17982249; GSPDB:GN00190  
 A:Experimental source: strain 16M  
 C:Genetics:  
 A:Gene: BME10351  
 A:Map position: 1  
 C:Superfamily: adenylosuccinate synthase  
 C:Keywords: ligase

Query Match 2.7%; Score 7; DB 2; Length 520;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 GYVLGIT 176  
 |||||  
 Db 350 GYVLGIT 356

RESULT 56  
 WHH0V4  
 tyrosine 3-monooxygenase (EC 1.14.16.2), splice form 4 - human  
 N:Alternate names: tyrosine 3-hydroxylase  
 N:Contains: tyrosine 3-monooxygenase, splice form 1; tyrosine 3-monooxygenase, splice  
 C:Species: Homo sapiens (man)  
 C>Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 16-Jun-2000  
 C:Accession: A30002; A26825; A60201; J00012; J00013; J00014; A27791; B27791; C27791; F  
 R:Nagatsu, T.  
 submitted to GenBank, December 1987  
 A:Reference number: A94509  
 A:Accession: A30002  
 A:Molecule type: mRNA  
 A:Residues: 1-528 <NAG1>  
 A:Cross-references: GB:M17589; NID:G339680; PIDN:AAA61179.1; PID:G339681  
 R:Kameda, N.; Kobayashi, K.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fuj  
 Blochem. Biophys. Res. Commun. 146, 971-975, 1987  
 A:Title: Isolation of a novel cDNA clone for human tyrosine hydroxylase: alternative F  
 A:Reference number: A30136; MUID:87298614; PMID:2887169  
 A:Accession: A26825  
 A:Molecule type: mRNA  
 A:Residues: 1-94 <NAG2>  
 A:Cross-references: GB:M17589; NID:G339680; PIDN:AAA61179.1; PID:G339681

R;Le Bourdellies, B.; Boularand, S.; Boni, C.; Horellou, P.; Dumas, S.; Grima, B.; Mallet  
 J. Neurochem. 50, 988-991, 1988  
 A;Title: Analysis of the 5' region of the human tyrosine hydroxylase gene: combinatorial  
 A;Reference number: A60201; MUID:88117543; PMID:2892893  
 A;Accession: A60201  
 A;Molecule type: mRNA  
 A;Residues: 1-65 <LEB>  
 A;Cross-references: GB:M4790; NID:g556223; PIDN:AAA61174.1; PID:g556224  
 R;Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fujita  
 U. Biochem. 103, 907-912, 1988  
 A;Title: Structure of the human tyrosine hydroxylase gene: alternative splicing from a s  
 A;Reference number: JE0012; MUID:89080200; PMID:2902075  
 A;Accession: JE0012  
 A;Molecule type: DNA  
 A;Residues: 1-30, 62-135 <KOB>  
 A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25094.1; PID:g2951764  
 A;Experimental source: splice form 1  
 A;Note: this splice form is produced by an alternative donor site within exon 1  
 A;Accession: JE0013  
 A;Molecule type: DNA  
 A;Residues: 1-34, 62-135 <KOB>  
 A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25097.1; PID:g2951767  
 A;Experimental source: splice form 2  
 A;Accession: JE0014  
 A;Molecule type: DNA  
 A;Residues: 1-30, 35-135 <KOB>  
 A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25095.1; PID:g2951765  
 A;Experimental source: splice form 3  
 A;Note: this splice form is produced by an alternative donor site within exon 1  
 R;Grima, B.; Lamouroux, A.; Boni, C.; Julien, J.F.; Javoy-Agid, F.; Mallet, J.  
 Nature 326, 707-711, 1987  
 A;Title: A single human gene encoding multiple tyrosine hydroxylases with different pred  
 A;Reference number: A93393; MUID:87173064; PMID:2882428  
 A;Accession: A27791  
 A;Molecule type: mRNA  
 A;Residues: 1-30, 62-528 <GR1>  
 A;Cross-references: GB:X05290; NID:g32501; PIDN:CAA28908.1; PID:g32502  
 A;Experimental source: splice form 1  
 A;Note: this splice form is produced by an alternative donor site within exon 1  
 A;Accession: B27791  
 A;Molecule type: mRNA  
 A;Residues: 1-34, 62-528 <GR1>  
 A;Cross-references: GB:X05290; NID:g32501  
 A;Experimental source: splice form 2  
 A;Accession: C27791  
 A;Molecule type: mRNA  
 A;Residues: 30, 35-528 <GR1>  
 A;Cross-references: GB:X05290; NID:g32501  
 A;Experimental source: splice form 3  
 A;Note: this isozyyme is produced by use of an alternative donor site within exon 1  
 R;Ichinose, H.; Onye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.  
 Biochem. Biophys. Res. Commun. 195, 158-165, 1993  
 A;Title: Increased heterogeneity of tyrosine hydroxylase in humans.  
 A;Reference number: PN0575; MUID:93371398; PMID:7689834  
 A;Accession: PN0575  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 19-30 <ICH1>  
 A;Accession: PN0582  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 35-61 <ICH2>  
 A;Accession: PN0588  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 62-106 <ICH3>  
 R;O'Malley, K.L.; Abhalt, M.J.; Martin, B.M.; Keisoe, J.R.; Winfield, S.L.; Gims, E.I.  
 Biochemistry 26, 2910-2914, 1987  
 A;Title: Isolation and characterization of the human tyrosine hydroxylase gene: identifi  
 A;Reference number: I52396; MUID:88107612; PMID:2892528  
 A;Accession: I52396  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: DNA

A;Residues: 1-61 <OMA>  
 A;Cross-references: GB:ML8116; NID:g339633; PIDN:AAA77649.1; PID:g1004335  
 R;Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fuj  
 Nucleic Acids Res. 15, 6733, 1987  
 A;Title: Isolation of a full-length cDNA clone encoding human tyrosine hydroxylase ty  
 A;Reference number: I38340; MUID:87316931; PMID:2888085  
 A;Accession: I38340  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-30, 35-528 <KOB4>  
 A;Cross-references: EMBL:Y00414; NID:g37126; PIDN:CAA68472.1; PID:g37127  
 R;Gims, E.I.; Rehavi, M.; Martin, B.M.; Weller, M.; O'Malley, K.L.; Lamarca, M.E.; M  
 U. Biol. Chem. 263, 7406-7410, 1988  
 A;Title: Expression of human tyrosine hydroxylase cDNA in invertebrate cells using a  
 A;Reference number: I55282; MUID:88213428; PMID:2896667  
 A;Accession: I55282  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-30, 62-64 <GIN1>  
 A;Cross-references: GB:M20911; NID:g339636; PIDN:AAA61167.1; PID:g339637  
 A;Accession: I70056  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-34, 62-64 <GIN2>  
 A;Cross-references: GB:M20912; NID:g339642; PIDN:AAA61168.1; PID:g339643  
 C;Comment: The expression of the four distinct proteins produced by alternate splicing  
 C;Genetics:  
 A;Gene: GDB:TH  
 A;Cross-references: GDB:119612; OMIM:191290  
 A;Map position: 11p15.5-11p15.5  
 A;Introns: 34/3; 61/3; 135/3  
 A;Note: the list of introns is incomplete  
 C;Function:  
 A;Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylala  
 A;Pathway: catecholamine biosynthesis  
 C;Note: this is the rate-limiting step in catecholamine biosynthesis  
 C;Specificity: phenylalanine 4-monooxygenase  
 C;Keywords: alternative splicing; biotin; catecholamine biosynthesis; iron; metal1  
 F;1-528/Product: tyrosine 3-monooxygenase, splice form 4 #status predicted <MRT4>  
 F;1-34, 62-528/Product: tyrosine 3-monooxygenase, splice form 2 #status predicted <MRT  
 F;1-30, 35-528/Product: tyrosine 3-monooxygenase, splice form 3 #status predicted <MRT  
 F;1-30, 62-528/Product: tyrosine 3-monooxygenase, splice form 1 #status predicted <MRT  
 F;8/Binding site: phosphate (Thr) (covalent) (by unidentified kinase) #status predict  
 F;19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #statu  
 F;361,366,406/Binding site: iron (His, His, Glu) #status predicted

Query Match 2.7%; Score 7; DB 1; Length 528;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	142 EAAAYVP 148	Db	446 EAAAYVP 452

RESULT 57  
 H86777  
 hypothetical protein ilvB [imported] - Lactococcus lactis subsp. lactis (strain IL140  
 C;Species: Lactococcus lactis subsp. lactis  
 C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #ext\_change 03-Aug-2001  
 C;Accession: H86777  
 R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Et  
 Genome Res. 11, 731-753, 2001  
 A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
 A;Reference number: A86825; MUID:21235186; PMID:11337471  
 A;Accession: H86777  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-540 <STO>  
 A;Cross-references: GB:AE005176; PID:g12724194; PIDN:AAK05322.1; GSPDB:GN00146  
 A;Experimental source: strain IL1403  
 C;Genetics:

A:Gene: 11VB  
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain h

Query Match 2.7%; Score 7; DB 2; Length 540;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LLAAYG 21  
DB 469 LLAAYG 475

## RESULT 58

D43719

urease (EC 3.5.1.5) 62K chain - Proteus mirabilis  
N:Alternate names: urease alpha chain

C:Species: Proteus mirabilis  
C>Date: 03-Mar-1993 #sequence\_revision 02-Dec-1994 #ext\_change 05-Jan-2003

C:Accession: D43719  
R:Jones, B.D.; Mobley, H.L.T.

J. Bacteriol. 171, 6414-6422, 1989

A:Title: Proteus mirabilis urease: nucleotide sequence determination and comparison with  
A:Reference number: A43719; MUID:90078080; PMID:268723

A:Accession: D43719  
A:Molecule type: DNA

A:Residues: 1-567 <CON>  
A:Cross-references: GB:M31834; NID:g150914; PIDN:AAA25669.1; PID:g150918

C:Genetics:  
A:Gene: urec

C:Complex: trimer of heterotrimeric subunits consisting of 11K (gamma), 12K (beta), and  
C:Function:

A:Description: catalyzes the hydrolysis of one molecule of urea to carbon dioxide and tw  
C:Superfamily: urease, alpha subunit; urease 62K chain homology

C:Keywords: heterotrimer; hydrolase; metalloprotein; nickel  
F:4-550/Domain: urease 62K chain homology <U62>

F:134.136.217.360/Binding site: nickel 2 (His, His, Lys, Asp) #status predicted  
F:217.246.272/Binding site: nickel 1 (Lys, His, His) #status predicted

F:217/Binding site: carbon dioxide (Lys) (covalent) (by urease activase) #status predict  
F:219.320/Active site: His #status predicted

Query Match 2.7%; Score 7; DB 1; Length 567;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 EAGVPEK 93  
DB 498 EAGVPEK 504

## RESULT 59

S35138

acetolactate synthase (EC 4.1.3.18) - Lactococcus lactis subsp. lactis  
C:Species: Lactococcus lactis subsp. lactis

A:Variety: strain NCD02118  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #ext\_change 28-Jul-2000

C:Accession: S35138  
R:Godon, J.J.; Chopin, M.C.; Ehrlich, S.D.

J. Bacteriol. 174, 6580-6589, 1992  
A:Title: Branched-chain amino acid biosynthesis genes in Lactococcus lactis subsp. lacti

A:Reference number: S35132; MUID:93015710; PMID:1400210  
A:Accession: S35138

A:Molecule type: DNA  
A:Residues: 1-575 <GOD>  
A:Cross-references: EMBL:M90761; NID:g2565137; PIDN:AAB81919.1; PID:g2565157

C:Genetics:  
A:Gene: 11VB

C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain h  
C:Keywords: branched-chain amino acid biosynthesis; carbon-carbon lyase; flavoprotein; m

F:45-483/Domain: thiamin pyrophosphate-binding domain homology <TPB>  
F:51/Active site: Glu #status predicted

Query Match 2.7%; Score 7; DB 1; Length 575;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LLAAYG 21  
DB 504 LLAAYG 510

## RESULT 60

S74477

hypothetical protein slr1114 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.  
A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #ext\_change 20-Jun-2000  
C:Accession: S74477

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

s.  
A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S74477  
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-585 <KAN>

A:Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BA16629.1; PID:g165  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Synechocystis hypothetical protein slr1114

Query Match 2.7%; Score 7; DB 2; Length 585;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 IAIAGI 188  
DB 512 IAIAGI 518

## RESULT 61

S77707

phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3) - yeast (Sacch

N:Alternate names: 5-aminomidazole-4-carboxamide ribotide transformylase; protein L1  
C:Species: Saccharomyces cerevisiae

C>Date: 02-May-1997 #sequence\_revision 02-May-1997 #ext\_change 18-Jun-1999  
C:Accession: S77707; S64855

R:Ribbets, A.S.; Appling, D.R.  
submitted to the EMBL Data Library, June 1996

A:Description: Isolation and characterization of two yeast genes encoding 5-aminomide  
A:Reference number: S77707

A:Accession: S77707  
A:Molecule type: DNA  
A:Residues: 1-591 <TRB>

A:Cross-references: EMBL:U62402; NID:g1480727; PIDN:AAB57774.1; PID:g1480728  
R:Obermaier, B.; Piravandi, E.; Rinke, M.

submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64845

A:Accession: S64855  
A:Molecule type: DNA

A:Residues: 1-545, 'TRILEC', 553, 'L', 555 <OBB>  
A:Cross-references: EMBL:Z79200; MIPS:YLR028c

A:Experimental source: strain S288c  
A>Note: this sequence has been revised in reference S77707

C:Genetics:  
A:Gene: SGD:ADE16

A:Cross-references: SGD:S0004018; MIPS:YLR028c  
A:Map position: 12R

C:Superfamily: purH bifunctional enzyme  
C:Keywords: hydrolase; purine nucleotide biosynthesis; transferase

Query Match 2.7%; Score 7; DB 2; Length 591;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 KDLKEQH 204

Db 87 KDLKEQH 93

# RESULT 62

SS4489

Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3) - yeast (Saccharomyces cerevisiae)

C/Species: Saccharomyces cerevisiae

C/Date: 08-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 05-Nov-1999

C/Accession: S54489

R/Lye, G.; Churcher, C.M.

Submitted to the EMBL Data Library, May 1995

A/Reference number: S54014

A/Accession: S54489

A/Molecule type: DNA

A/Residues: 1-592 <LYE>

A/Cross-references: EMBL:Z49273; NID:G809577; PIDN:CAA89269.1; PID:G809579; GSPDB:GN0001

C/Genetics:

A/Gene: SGD:ADE17; MIPS:YMR120C

A/Cross-references: SGD:S0004727; MIPS:YMR120C

A/Map position: 13R

C/Superfamily: purH bifunctional enzyme

C/Keywords: hydrolase; purine nucleotide biosynthesis; transferase

Query Match 2.7%; Score 7; DB 2; Length 592;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 KDLKEQH 204  
Db 87 KDLKEQH 93

# RESULT 63

C87048

Probable ABC transporter, ATP-binding component ML113 [imported] - Mycobacterium leprae

C/Species: Mycobacterium leprae

C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C/Accession: C87048

R/Cole, S.T.; Elgimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hume, M.A.; Rutherford, K.M.

Submitted to the EMBL Data Library, May 1995

A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sc

A/Title: Massive gene decay in the leprosy bacillus.

A/Reference number: A86909; MUID:21128732; PMID:11234002

A/Accession: C87048

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-629 <STO>

A/Cross-references: GB:AL450380; NID:G13093097; PIDN:CAC31494.1; GSPDB:GN00147

C/Genetics:

A/Gene: ML113

C/Superfamily: Escherichia coli ABC transporter mdIA; ATP-binding cassette homology

Query Match 2.7%; Score 7; DB 2; Length 629;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 QSGGLASA 59  
Db 355 QSGGLASA 361

QY 53 QSGGLASA 59  
Db 355 QSGGLASA 361

# RESULT 64

T15370

hypothetical protein CO1F1.3 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T15370

R/Johnson, D.

Submitted to the EMBL Data Library, May 1996

A/Description: The sequence of C. elegans cosmid CO1F1.

A/Reference number: Z18338

A/Accession: T15370

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-631 <JOH>

A/Cross-references: EMBL:U8761; NID:G1330391; PID:G1330395; PIDN:AAB00715.1; GSPDB:G

A/Experimental source: strain Bristol N2; clone CO1F1

C/Genetics:

A/Gene: CESP:CO1F1.3

A/Map position: 2

A/Introns: 80/3; 375/3; 423/2; 464/1; 490/3

Query Match 2.7%; Score 7; DB 2; Length 631;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 RITPLPS 221  
Db 498 RITPLPS 504

# RESULT 65

C97910

ATP-dependent proteinase ATP-binding chain [imported] - Streptococcus pneumoniae (strain

C/Species: Streptococcus pneumoniae

C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 17-May-2002

C/Accession: C97910

R/Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S

y, P.; Sun, P.M.; Winkler, M.E.

Submitted to the EMBL Data Library, May 1995

A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S

A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain Re.

A/Reference number: A97872; MUID:21429245; PMID:11544234

A/Accession: C97910

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-701 <KUR>

A/Cross-references: GB:AE007317; PIDN:AAK99111.1; PID:G15457861; GSPDB:GN00174

C/Genetics:

A/Gene: cplP

C/Superfamily: endopeptidase Clp ATP-binding chain

Query Match 2.7%; Score 7; DB 2; Length 701;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 EIOEASE 121  
Db 101 EIOEASE 107

# RESULT 66

I51283

hepatocyte growth factor precursor - clawed frog

N/Alternate names: hepatoleitin A; scatter factor

C/Species: Xenopus sp. (clawed frog)

C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 18-Jun-1999

C/Accession: I51283

R/Nakamura, H.; Tashiro, K.; Nakamura, T.; Shiokawa, K.

Submitted to the EMBL Data Library, May 1996

A/Title: Molecular cloning of Xenopus HGF cDNA and its expression studies in Xenopus

A/Reference number: I51283; MUID:95267690; PMID:7748783

A/Accession: I51283

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-710 <NAX>

A/Cross-references: GB:S77422; NID:G998932; PIDN:AA834354.1; PID:G998933

A/Note: the authors' translation for residue 458 (Thr) is inconsistent with the nucle

C/Complex: disulfide-bonded heterodimer of chains derived from the same precursor

C/Function: stimulates mitosis of hepatocytes and other cells

A/Description: stimulates mitosis of hepatocytes and other cells

A>Note: does not have proteinase activity  
C:Superfamily: hepatocyte growth factor; kringlike homology; trypsin homology  
C:Keywords: duplication; glycoprotein; growth factor; heterodimer; kringlike  
F:42-477,478-709/Product: hepatocyte growth factor #status predicted <MAT>  
F:42-477/Domains: hepatocyte growth factor alpha chain #status predicted <ACH>  
F:115-193/Domains: kringlike homology <KR1>  
F:198-275/Domains: kringlike homology <KR2>  
F:289-367/Domains: kringlike homology <KR3>  
F:375-453/Domains: kringlike homology <KR4>  
F:478-709/Domains: hepatocyte growth factor beta chain #status predicted <BCH>  
F:478-700/Domains: trypsin homology <TRY>  
F:52,128,281,322,379,550,637,666/Binding site: carbonylde (Asn) (covalent) #status pre  
F:470-588/Diulfide bonds: #status predicted

Query Match 2.7%; Score 7; DB 1; Length 710;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

196 RGDLMKE 202  
155 RGDLMKE 161

RESULT 67  
T38359  
cullin 3 homolog - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T38359  
R:Skellton, J.; Churcher, C.M.; Barrell, B.G.; Rajindream, M.A.; Walsh, S.V.; Wood, V.  
submitted to the EMBL Data Library, September 1995  
A:Reference number: Z21788  
A:Accession: T38359  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-785 <SKE>  
A:Cross-references: EMBL:Z54142; PDB:CAA9847.1; GSPDB:GN00066; SPDB:SPAC24H6.03  
A:Experimental source: strain 972h-; cosmid c24H6  
C:Genetics:  
A:Gene: pcu3; SPDB:SPAC24H6.03  
A:Map position: 1  
A:Introns: 14/3; 476/2; 513/1; 534/3; 592/3; 729/2

Query Match 2.7%; Score 7; DB 2; Length 785;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

103 ETTSQL 109  
353 ETTSQL 359

RESULT 68  
S62405  
hypothetical protein SPAC24H6.03 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 31-Oct-1997  
C:Accession: S62405  
R:Skellton, J.; Churcher, C.M.  
submitted to the EMBL Data Library, September 1995  
A:Reference number: S62402  
A:Accession: S62405  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-798 <SKR>  
A:Cross-references: EMBL:Z54142; NID:g984697; PID:g984700  
C:Genetics:  
A:Map position: 1L  
A:Introns: 14/3; 526/1; 547/3; 605/3; 742/2

Query Match 2.7%; Score 7; DB 2; Length 798;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 ETTSQL 109  
DB 353 ETTSQL 359

RESULT 69  
A45082  
neurotrophic receptor rorl precursor - human  
N:Contains: protein-tyrosine kinase (BC 2.7.1.112)  
C:Species: Homo sapiens (man)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-May-2000  
C:Accession: A45082  
R:Maslowski, P.; Carroll, R.D.  
J. Biol. Chem. 267, 26181-26190, 1992  
A:Title: A novel family of cell surface receptors with tyrosine kinase-like domain.  
A:Reference number: A45082; MUID:93100347; PMID:1334494  
A:Accession: A45082  
A:Molecule type: mRNA  
A:Residues: 1-937 <MAS>  
A:Cross-references: GB:M97675; NID:9337464; PID:AAA60275.1; PID:9337465  
A>Note: sequence extracted from NCBI backbone (NCBIP:120916)  
C:Genetics:  
A:Gene: GDB:NTKRK1  
A:Cross-references: GDB:136453  
A:Map position: 6p21-6p21  
C:Superfamily: neurotrophic receptor rorl; immunoglobulin homology; kringlike homology;  
C:Keywords: ATP; glycoprotein; kringlike; phosphotransferase; transmembrane protein; ty  
F:1-23/Domains: signal sequence #status predicted <SIG>  
F:124-937/Product: neurotrophic receptor rorl #status predicted <MAT>  
F:12-133/Domains: immunoglobulin homology <IMM>  
F:313-391/Domains: kringlike homology <KR>  
F:404-425/Domains: transmembrane #status predicted <TM1>  
F:471-753/Domains: protein kinase homology <KIN>  
F:47,66,184,315/Binding site: carbonylde (Asn) (covalent) #status predicted

Query Match 2.7%; Score 7; DB 2; Length 937;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

67 HSYCRNP 73  
359 HSYCRNP 365

RESULT 70  
G70951  
Probable ATP-dependent DNA helicase - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: G70951  
R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
i Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Najdareem, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squires, S.  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Skellton, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: G70951  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1101 <COL>  
A:Cross-references: GB:AL021646; GB:AL123456; NID:93242278; PID:CAA16666.1; PID:e1246  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV3201c

Query Match 2.7%; Score 7; DB 2; Length 1101;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

57 ASAPVSG 63

Db 594 ASAPVSG 600

RESULT 71  
S49247  
parasporal crystal protein cry9cal [validated] - *Bacillus thuringiensis*  
N/Alternate names: parasporal crystal protein cryIH  
C/Species: *Bacillus thuringiensis*  
C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 01-Dec-2000  
C/Accession: A59350, S49247  
R/Lambert, B.; Buysse, L.; Decock, C.; Janssens, S.; Pien, C.; Seey, B.; Seurinck, J.; V  
Appel, Environ. Microbiol. 62, 80-86, 1996  
A/Title: A *Bacillus thuringiensis* insecticidal crystal protein with a high activity aga  
A/Reference number: A59350; MUID:96141404; PMID:8572715  
A/Accession: A59350  
A/Molecule type: DNA  
A/Residues: 1-1157 <L&M>  
A/Cross-references: EMBL:Z37527; NID:9547554; PIDN:CAA5764.1; PID:9547556  
A/Experimental source: serovar tolworthi  
C/Comment: This parasporal crystal protein, active against corn borer and other insects,  
C/Superfamily: parasporal crystal protein  
C/Keywords: delta-endotoxin

Query Match 2.7%; Score 7; DB 1; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ASAPVSG 63  
|||||  
Db 523 ASAPVSG 529

RESULT 72  
T16859  
hypothetical protein T13C2.5 - *Caenorhabditis elegans*  
C/Species: *Caenorhabditis elegans*  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C/Accession: T16859  
R/Du, Z.  
submitted to the EMBL Data Library, November 1995  
A/Description: The sequence of C. elegans coemid T13C2.  
A/Reference number: Z16591  
A/Accession: T16859  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1296 <DUZ>  
A/Cross-references: EMBL:U40030; NID:91055164; PID:91055165; PIDN:AAA81133.1; CESP:T13C2  
C/Genetics:  
A/Gene: CESP:T13C2.5  
A/Introns: 62/3; 93/3; 124/3; 142/1; 212/1; 293/1; 336/2; 406/3; 437/1; 508/1; 584/1; 62

Query Match 2.7%; Score 7; DB 2; Length 1296;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 PCEDLRC 101  
|||||  
Db 216 PCEDLRC 222

RESULT 73  
H70846  
hypothetical glycine-rich protein RV3345c - *Mycobacterium tuberculosis* (strain H37RV)  
C/Species: *Mycobacterium tuberculosis*  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C/Accession: H70846  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S  
; Connor, R.; Davies, R.; Devlin, K.; Fieldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A/Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A/Reference number: A70500; MUID:96293987; PMID:9634230

A/Accession: H70846  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-1538 <COL>  
A/Cross-references: GB:AL021841; GB:AL123456; NID:93261517; PIDN:CAA17117.1; PID:9326  
C/Genetics:  
A/Gene: RV3345c  
C/Superfamily: collagen alpha 1 (IV) chain

Query Match 2.7%; Score 7; DB 2; Length 1538;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 GOAGTPG 262  
|||||  
Db 776 GOAGTPG 782

RESULT 74  
T30313  
chemotaxis protein homolog - *Pseudomonas aeruginosa*  
C/Species: *Pseudomonas aeruginosa*  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C/Accession: T30313  
R/Whitchurch, C.B.; Young, M.D.; Hobbs, M.; Mattick, J.S.  
submitted to the EMBL Data Library, November 1996  
A/Description: *Pseudomonas aeruginosa* chemotactic transduction genes pilL, chpA chpB  
A/Reference number: Z20819  
A/Accession: T30313  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1638 <WHI>  
A/Cross-references: EMBL:U79580; NID:93241967; PID:93241969; PIDN:AAC2931.1  
C/Genetics:  
A/Note: chpA

Query Match 2.7%; Score 7; DB 2; Length 1638;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 LDAQSGI 56  
|||||  
Db 919 LDAQSGI 925

RESULT 75  
E83594  
still frameshift probable component of chemotactic signal transduction system PA0413  
C/Species: *Pseudomonas aeruginosa*  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/Accession: E83594  
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; I  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; I  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa  
A/Reference number: A82950; MUID:20473337; PMID:10984043  
A/Accession: E83594  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-2472 <STO>  
A/Cross-references: GB:AE004478; GB:AE004091; NID:99946261; PIDN:AA03802.1; GSPDB:GN  
C/Experimental source: strain PA01  
C/Genetics:  
A/Gene: PA0413

Query Match 2.7%; Score 7; DB 2; Length 2472;  
Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 LDAQSGI 56  
|||||



Db 1753 LDAOSGL 1759

Search completed: November 25, 2003, 13:41:42  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:34:29 ; Search time 18 Seconds

(without alignments)  
687.112 Million cell updates/sec

Title: US-10-057-951-2

Perfect score: 263  
Sequence: 1 MLTAWQAEFLVSNMLAEAY.....PVDPGSGSTPMGQAGTPGA 263

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 127863 segs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	3.0	429	PUR4_THIFB	P52151 chibacillin
2	8	3.0	716	HGFL_MOUSE	P26928 mus musculu
3	8	3.0	771	PCRA_MYCTU	P71561 mycobacteri
4	7	2.7	96	CH10_AMOPS	P26005 amocba proc
5	7	2.7	144	GLB3_LAMSP	P15469 lamedlibrac
6	7	2.7	144	GLB8_RIFPA	P80592 rifitla pach
7	7	2.7	154	YL66_ARCFU	O28116 archaeoglob
8	7	2.7	229	PRL_MELGA	P17572 meleagris g
9	7	2.7	281	STO4_CAEBL	O22165 caenorhabd
10	7	2.7	301	SC14_CANAL	P46250 candida alb
11	7	2.7	335	Y780_METUA	P43798 haemophilus
12	7	2.7	335	Y780_METUA	Q58190 methanococc
13	7	2.7	351	ARGC_SYNY3	P54899 synchocyst
14	7	2.7	388	XYLR_LACPE	P21940 lactobacill
15	7	2.7	398	AAT_PSEAE	P72173 pseudomonas
16	7	2.7	402	PUR4_BRUAB	P52004 bruceella ab
17	7	2.7	429	PUR4_BRUAB	Q8y1ul bruceella me
18	7	2.7	432	PUR4_RHIL0	Q98197 rhizobium 1
19	7	2.7	456	FTSA_PORGI	O07827 porphyromon
20	7	2.7	512	EMRB_ECOLI	P27304 escherichia
21	7	2.7	528	TY3H_HUMAN	P07101 homo sapien
22	7	2.7	567	URR1_PROMI	P17086 proteus mir
23	7	2.7	575	ILVB_LACLA	Q02137 lactococcus
24	7	2.7	591	PUR1_YEAST	P54113 s bifunctio
25	7	2.7	592	PUR2_YEAST	P38009 s bifunctio
26	7	2.7	619	LT65_ARATH	Q04880 arabidopsis
27	7	2.7	661	UAS3_HUMAN	P57075 homo sapien
28	7	2.7	769	LEM3_SHEEP	P98109 ovis aries
29	7	2.7	785	CUT3_SCHRO	Q09760 schizosacch
30	7	2.7	937	ROR1_HUMAN	Q01973 homo sapien
31	7	2.7	937	ROR1_MOUSE	Q92139 mus musculu
32	7	2.7	1034	ADD_DROME	P54362 drosophila
33	7	2.7	1157	C9CA_BACTO	Q45733 bacillus th

34	7	2.7	3433	1	POLG_KUNJM	P14335 k genome po
35	6	2.3	41	1	LAMA_EMENT	P38095 emeticella
36	6	2.3	49	1	Y185_METUA	Q57644 methanococ
37	6	2.3	63	1	BD02_RAT	O88514 rattus norv
38	6	2.3	69	1	CSPA_PSEAE	P95459 pseudomonas
39	6	2.3	78	1	PYS1_SYNEL	P50035 synchocyst
40	6	2.3	81	1	ATP1_MYCLE	P45828 mycobacteri
41	6	2.3	81	1	ATP1_MYCTU	Q10598 mycobacteri
42	6	2.3	83	1	RS17_CHLMU	Q9pjln3 chlamydia m
43	6	2.3	83	1	RS17_CHLTR	P28545 chlamydia t
44	6	2.3	86	1	RS17_CHLTP	Q92726 chlamydia t
45	6	2.3	93	1	IHEP_PASHA	P95519 pasteurella
46	6	2.3	95	1	Y492_METUA	O57915 methanococ
47	6	2.3	105	1	THIO_ALIAC	P80579 alicyclobac
48	6	2.3	106	1	RLA4_YEAST	P10622 saccharomyc
49	6	2.3	106	1	THCC_RHOER	P43493 rhodococcus
50	6	2.3	108	1	DBH_BORAD	O44625 borrelia an
51	6	2.3	108	1	DBH_BORAF	O57220 borrelia af
52	6	2.3	108	1	DBH_BORBU	O57267 borrelia bu
53	6	2.3	108	1	DBH_BORJA	O57153 borrelia ga
54	6	2.3	108	1	DBH_BORTR	O45722 borrelia ja
55	6	2.3	109	1	RLA1_SCHPO	O45722 borrelia tu
56	6	2.3	109	1	RLA5_SCHPO	P17476 schizosacch
57	6	2.3	109	1	RS17_RALNY	Q9uu78 schizosacch
58	6	2.3	109	1	RS17_RALNY	O24786 halobacteri
59	6	2.3	110	1	RLA3_SCHPO	P17477 schizosacch
60	6	2.3	111	1	RS17_HALMA	P12741 haloarcula
61	6	2.3	118	1	IAA4_SORBI	P81367 sorghum bic
62	6	2.3	121	1	YHAA_ECOLI	P42651 escherichia
63	6	2.3	122	1	RL7_LIBAS	P36247 libibactate
64	6	2.3	124	1	NOS_VICSA	Q41705 vicia sativ
65	6	2.3	130	1	CAU2_MOUSE	Q99mp3 mus musculu
66	6	2.3	133	1	RNPA_CORGL	O8n151 corynebacte
67	6	2.3	135	1	H32_TERYP	P15512 tetrahymena
68	6	2.3	135	1	H33_TERTH	P41353 tetrahymena
69	6	2.3	135	1	NOS_PBA	P25226 pisum sativ
70	6	2.3	136	1	YPO9_DEIRA	Q9xrl2 haemophilus
71	6	2.3	139	1	YQGF_HAEIN	P43981 haemophilus
72	6	2.3	141	1	HBA2_TRICR	P10784 triturus cr
73	6	2.3	144	1	RIB1_PHOLE	O01994 photobacter
74	6	2.3	145	1	DSBH_OCEIH	O8ey3 oceanobacil
75	6	2.3	148	1	Y222_TREPA	P08097 escherichia
76	6	2.3	148	1	Y222_TREPA	O8351 treponema p
77	6	2.3	150	1	YQK_ECOLI	P39335 escherichia
78	6	2.3	153	1	SODC_NEUCR	P07509 neurospora
79	6	2.3	156	1	SPRT_HAEIN	P44119 haemophilus
80	6	2.3	160	1	YQGC_BACSU	P54486 bacillus su
81	6	2.3	169	1	PLMN_RAT	O01177 rattus norv
82	6	2.3	173	1	LEPA_MYCHY	Q9zn28 mycoplasma
83	6	2.3	175	1	PR18_STRCH	P77949 streptomyce
84	6	2.3	176	1	YV29_MYCTU	O50703 mycobacteri
85	6	2.3	178	1	YV29_MYCTU	O06246 mycobacteri
86	6	2.3	180	1	HELO_PHYPA	Q9m330 physcomitri
87	6	2.3	185	1	ADML_RAT	P43143 rattus norv
88	6	2.3	187	1	RPE1_METUA	Q57840 methanococ
89	6	2.3	192	1	NIFZ_FRAAL	P46040 frankia ain
90	6	2.3	193	1	R122_ARATH	P36211 arabidopsis
91	6	2.3	198	1	GGR2_HUMAN	P36268 homo sapien
92	6	2.3	201	1	Y802_NEIMA	O91k33 nessesetia m
93	6	2.3	204	1	PDGA_MOUSE	P14801 human immun
94	6	2.3	206	1	NEE_FVLEL	P04604 human immun
95	6	2.3	207	1	NEE_FVLEL	P04604 human immun
96	6	2.3	211	1	PDGA_HUMAN	P04085 homo sapien
97	6	2.3	211	1	PDGA_MOUSE	P20033 mus musculu
98	6	2.3	212	1	ALKH_HAEIN	P44480 h putative
99	6	2.3	214	1	PSAE_YERPE	P31224 yersinia pe
100	6	2.3	218	1	YXU1_BACSU	P55184 bacillus su

## ALIGNMENTS

RESULT 1

PUR1 TH1FE STANDARD; PRT: 429 AA.

AC P25151;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase) (AAS5) (AMPSase).

GN PUR1.

OS Thobacillus ferrooxidans.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;

OC Acidithiobacillaceae; Acidithiobacillus.

OX NCBI\_TaxID=920;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Fe1;

RX MEDLINE=92283830; PubMed=1317860;

RA Kubano T., Takeshima T., Sugawara K., Inoue C., Shiratori T., Yano T., Fukumori Y., Yamanaka T.;

RT "Molecular cloning of the gene encoding Thobacillus ferrooxidans Fe(II) oxidase. High homology of the gene product with HripP.";

RL J. Biol. Chem. 267:11242-11247(1992).

CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE NUCLEOTIDE BIOSYNTHESIS.

CC CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate + adenylosuccinate.

CC -1- PATHWAY: AMP biosynthesis; first committed step.

CC -1- SIMILARITY: Belongs to the adenylosuccinate synthetase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X57324; CAA0593.1; -.

DR PIR: S23258; S23258.

DR HSSP: P12283; IADE.

DR HAMAP: MF\_00011; -; 1.

DR InterPro: IPR001114; Asucc synthetase.

DR Pfam: PF00709; Adenylsuc. synt. 1.

DR ProDom: PD001188; Asucc\_synthetase; 1.

DR TIGRfam: TIGR00184; purA; 1.

DR PROSITE: PS00513; ADENYLOSUCIN SYN 2; 1.

DR PROSITE: PS01266; ADENYLOSUCIN SYN 1; 1.

KW Purine biosynthesis; ligase; GTP-binding.

FT NP BIND 13 19

FT ACT\_SITE 141 141 BY SIMILARITY.

FT ACT\_SITE 148 148

SQ SEQUENCE 429 AA; 46355 MW; C5B2F39C3E105719 CRC64;

Query Match 3.0%; Score 8; DB 1; Length 429;

Best Local Similarity 100.0%; Pred. No. 3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 LGYVLGIT 176

Db 260 LGYVLGIT 267

RESULT 2

HGFL\_MOUSE STANDARD; PRT: 716 AA.

AC P26928;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Hepatocyte growth factor-like protein precursor (Macrophage stimulatory protein) (MSP).

DE MS1 OR HGFL.

GN Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C; TISSUE=Liver;

RX MEDLINE=92002017; PubMed=1832957;

RA Fritzenr Degen S.D., Stuart L.A., Han S., Jamison C.S.;

RT "Characterization of the mouse cDNA and gene coding for a hepatocyte growth factor-like protein: expression during development.";

RL Biochemistry 30:9781-9791(1991).

CC -1- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA CHARACTERISTIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT CONSERVED.

CC -1- TISSUE SPECIFICITY: LIVER. LOWER LEVELS IN LUNG, PLACENTA AND ADRENAL.

CC -1- DEVELOPMENTAL STAGE: IS EXPRESSED AT LOW LEVELS DURING GESTATION. JUST BEFORE BIRTH THE LEVEL INCREASES DRAMATICALLY AND REMAINS STABLE AFTERWARDS.

CC -1- PTM: MAY BE CLEAVED AFTER AA 488, TO YIELD A TWO-CHAIN MOLECULE HELD TOGETHER BY DISULFIDE BONDS, OR CLEAVED INTO TWO SEPARATE POLYPEPTIDES.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMINOGEN SUBFAMILY.

CC -1- SIMILARITY: Contains 4 kringle domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M74180; AAA50166.1; -.

DR PIR: M74181; AAA50167.1; -.

DR PIR: A40332; A40332.

DR HSSP: P00747; IKRN.

DR MEROPS: S01.975; -.

DR MGD; MGI:96080; Mst1.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR000001; Kringle.

DR InterPro: IPR003014; PAN.

DR InterPro: IPR003609; Pan\_app.

DR InterPro: IPR003966; Prochrombin.

DR InterPro: IPR001254; Ser protease\_Try.

DR Pfam: PF00051; Kringle; 4.

DR Pfam: PF00024; PAN; 1.

DR Pfam: PF00089; trypsin; 1.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR PRINTS: PR00018; KRINGLE.

DR PRINTS: PR01505; PROTHROMBIN.

DR ProDom: PD000395; Kringle; 4.

DR SMART: SM00473; PAN\_AP; 1.

DR SMART: SM00473; PAN\_AP; 1.

DR SMART: SM00473; PAN\_AP; 1.

DR PROSITE: PS00020; Tryp\_Spc; 1.

DR PROSITE: PS00021; KRINGLE 1; 4.

DR PROSITE: PS00070; KRINGLE 2; 4.

DR PROSITE: PS00240; TRYPSIN\_DOM; 1.

KW Kringle; Glycoprotein; Serine protease homolog; Repeat; Signal.

FT SIGNAL 1 18

FT CHAIN 19 716

FT CHAIN 19 109

FT DOMAIN 110 186

FT DOMAIN 191 268

FT DOMAIN 292 370

FT DOMAIN 379 457

FT DOMAIN 489 716

FT DISULFID 56 78

FT DISULFID 60 66

FT DISULFID 110 186

FT DISULFID 131 169

FT DISULFID 157 181

FT DISULFID 191 268

BY SIMILARITY.

HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.

PAP.

KRINGLE 1.

KRINGLE 2.

KRINGLE 3.

KRINGLE 4.

SERINE PROTEASE-LIKE.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

```

FT DISULFID 194 333 INTERCHAIN (BY SIMILARITY).
FT DISULFID 212 251 BY SIMILARITY.
FT DISULFID 240 263 BY SIMILARITY.
FT DISULFID 292 370 BY SIMILARITY.
FT DISULFID 313 352 BY SIMILARITY.
FT DISULFID 341 364 BY SIMILARITY.
FT DISULFID 379 457 BY SIMILARITY.
FT DISULFID 400 440 BY SIMILARITY.
FT DISULFID 428 452 BY SIMILARITY.
FT DISULFID 477 593 INTERCHAIN (BY SIMILARITY).
FT DISULFID 512 528 BY SIMILARITY.
FT DISULFID 607 672 BY SIMILARITY.
FT DISULFID 637 651 BY SIMILARITY.
FT DISULFID 662 690 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 P -> Q (IN GENOMIC SEQUENCE).
SQ SEQUENCE 716 AA; 80588 MW; BBCE02EF85213ACC CRC64;

```

```

Query Match 3.0%; Score 8; DB 1; Length 716;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 76 DPGPCWCY 83
DB 163 DPGPCWCY 170

```

## RESULT 3

```

PCRA_MYCTU ID PCRA_MYCTU STANDARD; PRT; 771 AA.
AC P71561;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent DNA helicase pcrA (EC 3.6.1.-).
GN PCRA OR IVRD OR RV0949 OR MT0976 OR MTCY10D7.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekla F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sluskin J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., Derooy R., Dodson R., Gwin M.L., Hart D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z79700; CAB02001.1; -
DR EMBL; AE006982; AAK45224.1; -
DR PIR; C70716; C70716.
DR HSSP; P56255; IPR.
DR TIGR; MT0976; -
DR TubercuList; RV0949; -
DR InterPro; IPR005751; PCRA.
DR InterPro; IPR000212; UVRD-helicase.
DR Pfam; PF00580; UVRD-helicase; 1.
DR TIGRFAMs; TIGR01073; pcrA; 1.
KW DNA-binding; Hydrolase; Helicase; ATP-binding; DNA repair;
KW Complete proteome.
FT NP_BIND 42 49 ATP (POTENTIAL).
SQ SEQUENCE 771 AA; 85049 MW; DBAAB4E151FAE2C9 CRC64;

```

```

Query Match 3.0%; Score 8; DB 1; Length 771;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 58 SAPVSGAG 65
DB 692 SAPVSGAG 699

```

## RESULT 4

```

CH10_AMOPS ID CH10_AMOPS STANDARD; PRT; 96 AA.
AC P26005;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 10 kDa chaperonin (Protein Cpn10) (groES protein).
GN GROS OR GROES.
OS Amoeba proteus symbiotic bacterium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Legionellaceae.
OX NCBI_TaxId=2728;
RN [1]
RP SEQUENCE FROM N.A.
RA Ahn T.I., Lee H.K., Kwak I.H., Jeon K.W.;
RT "Nucleotide sequence and temperature-dependent expression of groEL
RT gene isolated from symbiotic bacteria of Amoeba proteus."
RL Endocyt. Cell Res. 8:33-44(1991).
CC -1- FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses
CC the ATPase activity of the latter.
CC -1- SUBUNIT: Heptamer of 7 subunits arranged in a ring (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86549; AAC09380.1; -
DR PIR; JC2561; JC2561.
DR HSSP; P05380; IAON.
DR HAMAP; MF 00580; -; 1.
DR InterPro; IPR001476; Chapernin_Cpn10.
DR Pfam; PF00166; cpn10; 1.
DR PRINTS; PR00297; CHAPERONIN10.
DR PRODOM; PD000566; Chapernin_Cpn10; 1.
DR PROSITE; PS00681; CHAPERONINS_CPN10; 1.

```

```

KW  Chaparone. 96 AA; 10618 MW; F6DEE1A8056B89B CRC64;
SQ  SEQUENCE

Query Match 2.7%; Score 7; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  181 IIAIGAG 187
    |||||
    40 IIAIGAG 46

RESULT 5
GLB3_LAMSP
ID  GLB3_LAMSP STANDARD; PRT; 144 AA.
AC  P15469;
DT  01-APR-1990 (Rel. 14, Created)
DT  01-APR-1990 (Rel. 14, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Giant hemoglobin AIII chain.
OS  Lamelibrachia sp. (Deep-sea giant tube worm).
OC  Eukaryota; Metazoa; Pogonophora; Vestimentifera; Basibranchia;
CX  Lamelibrachidae; Lamelibrachidae; Lamelibrachia.
NX  NCBI_TaxID=6424;
RN  [1]
RP  SEQUENCE.
RX  MEDLINE=90179711; PubMed=2310374;
RA  Suzuki T., Takagi T., Ohta S.;
RT  "Primary structure of a constituent polypeptide chain (AIII) of the
RT  giant haemoglobin from the deep-sea tube worm Lamelibrachia. A
RT  possible H2S-binding site.";
RL  Biochem. J. 266:221-225(1990).
RN  [2]
RP  SEQUENCE OF 1-24.
RX  MEDLINE=89076216; PubMed=3202832;
RA  Suzuki T., Takagi T., Ohta S.;
RT  "N-terminal amino acid sequence of the deep-sea tube worm haemoglobin
RT  remarkably resembles that of annelid haemoglobin.";
RL  Biochem. J. 255:541-545(1988).
CC  -1- SUBUNIT: GIANT HEMOGLOBIN IS COMPOSED OF FOUR HEME-CONTAINING
CC  CHAINS (AI TO AIV), AND TWO LINKER CHAINS (AV AND AVI).
CC  -1- SIMILARITY: Belongs to the globin family.
DR  PIR; S08284; S08284.
DR  InterPro; IPR002336; Erythrinin.
DR  InterPro; IPR000971; Globin.
DR  Pfam; PF00042; Globin; 1.
DR  PRINTS; PR00611; ERYTHROGLOBIN.
DR  PROSITE; PS01033; GLOBIN; 1.
KW  Heme; Oxygen transport; Transport.
SQ  SEQUENCE 144 AA; 16004 MW; 036CC2E9B1E7F69 CRC64;

Query Match 2.7%; Score 7; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  17 AEAYGSG 23
    |||||
    16 AEAYGSG 22

RESULT 6
GLB3_RIFPA
ID  GLB3_RIFPA STANDARD; PRT; 144 AA.
AC  P80592;
DT  15-JUL-1998 (Rel. 36, Created)
DT  15-JUL-1998 (Rel. 36, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Giant hemoglobins B chain.
OS  Riftia pachyptila (tube worm).
OC  Eukaryota; Metazoa; Pogonophora; Vestimentifera; Axonobranchia;
CX  Riftiida; Riftiidae; Riftia.
NX  NCBI_TaxID=6426;
RN  [1]

```

```

RP  SEQUENCE.
RX  MEDLINE=98073282; PubMed=9408952;
RA  Zai F., Suzuki T., Kawasaki Y., Childress J.J., Lallier F.H.,
RA  Toulmond A.;
RT  "Primary structure of the common polypeptide chain b from the multi-
RT  hemoglobin system of the hydrothermal vent tube worm Riftia
RT  pachyptila: an insight on the sulfide binding-site.";
RL  Proteins 29:562-574(1997).
CC  -1- SUBUNIT: PART OF GIANT HEMOGLOBIN C1, V1 AND V2. THIS WORM HAS
CC  THREE DIFFERENT EXTRACELLULAR HBS: TWO DISSOLVED IN THE VASCULAR
CC  BLOOD, V1 (CA. 3,500 KDa) AND V2 (CA. 400 KDa), AND ONE IN THE
CC  COELOMIC FLUID, C1 (CA. 400 KDa). V1 CONSISTS OF FOUR HEME-
CC  CONTAINING, GLOBIN CHAINS (B-E) AND FOUR LINKER CHAINS (L1-L4). V2
CC  CONSISTS OF SIX GLOBIN CHAINS (A-F) AND C1 CONSISTS OF FIVE GLOBIN
CC  CHAINS (A-E).
CC  -1- SUBCELLULAR LOCATION: Extracellular.
CC  -1- SIMILARITY: Belongs to the globin family.
DR  InterPro; IPR000971; Globin.
DR  Pfam; PF00042; Globin; 1.
DR  PROSITE; PS01033; GLOBIN; 1.
KW  Heme; Oxygen transport; Transport.
SQ  SEQUENCE 144 AA; 16135 MW; 9A094A9E8E981568 CRC64;

Query Match 2.7%; Score 7; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  17 AEAYGSG 23
    |||||
    17 AEAYGSG 23

RESULT 7
YL66_ARCFU
ID  YL66_ARCFU STANDARD; PRT; 154 AA.
AC  O28116;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Hypothetical protein AF2166.
GN  AF2166.
OS  Archaeoglobus fulgidus.
OC  Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC  Archaeoglobaceae; Archaeoglobus.
CX  NCBI_TaxID=2234;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX  MEDLINE=98049343; PubMed=9389475;
RA  Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA  Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA  Richardson D.L., Kerlavage A.R., Graham D.E., Kyprides N.C.,
RA  Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA  Kinkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA  Peterson S., Reich C.L., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA  Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uutterback T.,
RA  Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA  Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA  Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA  Venter J.C.;
RT  "The complete genome sequence of the hyperthermophilic, sulphate-
RT  reducing archaeon Archaeoglobus fulgidus";
RL  Nature 390:364-370(1997).
CC  -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----

```

DR EMBL; AE000955; AAB89089.1; -  
 DR PIR; F69520; F69520.  
 DR TIGR; AF2166; -  
 KM Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 15 37 POTENTIAL.  
 FT TRANSMEM 58 80 POTENTIAL.  
 FT TRANSMEM 95 116 POTENTIAL.  
 FT TRANSMEM 123 145 POTENTIAL.  
 SQ SEQUENCE 154 AA; 17266 MW; 1CE41361CD9FA68 CRC64;  
 Query Match 2.7%; Score 7; DB 1; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 187 GILIGYS 193  
 DB 104 GILIGYS 110  
 RESULT 8  
 PRL\_MELGA STANDARD; PRT; 229 AA.  
 AC P17572;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Prolectin precursor (PRL).  
 GN PRL.  
 OS Meleagris gallopavo (Common turkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.  
 OC NCBI\_Taxid=9103;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RX MEDLINE=96206340; PubMed=8618952;  
 RA Xu M., Proudman J.A., Pilts G.R., Wong E.A., Foster D.N.,  
 RA el Halawani M.E.;  
 RT "Vasoreactive intestinal peptide stimulates prolactin mRNA expression  
 RT in turkey pituitary cells: effects of dopaminergic drugs.";  
 RL Proc. Soc. Exp. Biol. Med. 212:52-62(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=91346480; PubMed=1879669;  
 RA Wong E.A., Ferrin N.H., Silsby J.L., el Halawani M.E.;  
 RT "Cloning of a turkey prolactin cDNA: expression of prolactin mRNA  
 RT throughout the reproductive cycle of the domestic turkey (Meleagris  
 RT gallopavo).";  
 RL Gen. Comp. Endocrinol. 83:18-26(1991).  
 RN [3]  
 RP SEQUENCE OF 66-229 FROM N.A.  
 RP TISSUE=Pituitary;  
 RX MEDLINE=90272435; PubMed=2349117;  
 RA Karatzas C.N., Zadworny D., Kuhnlein U.;  
 RT "Nucleotide sequence of turkey prolactin.";  
 RL Nucleic Acids Res. 18:3071-3071(1990).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U05957; AAB60615.1; -  
 DR EMBL; U05953; AAB60615.1; JOINED.  
 DR EMBL; U05954; AAB60615.1; JOINED.  
 DR EMBL; U05955; AAB60615.1; JOINED.  
 DR EMBL; U05952; AAB60604.1; -  
 DR EMBL; X51769; CAA36071.1; -  
 DR PIR; A61133; A61133.

DR HSSP; Q28632; 1AN3  
 DR InterPro; IPR001400; Somatotropin.  
 DR Pfam; PF00103; hormone; 1.  
 DR PRINTS; PR00836; SOMATOTROPIN.  
 DR PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
 KM Hormone; pituitary; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 229  
 FT DISULFID 34 41  
 FT DISULFID 88 204  
 FT DISULFID 221 229  
 FT CONFLICT 156 156  
 SQ SEQUENCE 229 AA; 25854 MW; DE4530EB2301F2B7 CRC64;  
 Query Match 2.7%; Score 7; DB 1; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 LVSNMML 16  
 DB 17 LVSNMML 23  
 RESULT 9  
 ST04\_CAEEL STANDARD; PRT; 281 AA.  
 AC Q22165; Q9N644;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Stomatatin protein 4.  
 GN ST0-4 OR Y71H9A.3/T04F8.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OC NCBI\_Taxid=6239;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RX STRAIN=Bristol N2;  
 RA Leonard N., White S.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP REVISIONS.  
 RP Durbin R.;  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE BAND 7 / MEC-2 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AL032659; CAB82215.1; -  
 DR EMBL; Z65565; CAB82215.1; JOINED.  
 DR EMBL; Z65565; CAB76415.1; -  
 DR EMBL; AL032659; CAB76415.1; JOINED.  
 DR WormPep; Y71H9A.3; CE22957.  
 DR InterPro; IPR001107; Band 7.  
 DR InterPro; IPR001972; Stomatatin.  
 DR Pfam; PF01145; Band 7; 1.  
 DR PRINTS; PR00721; STOMATIN.  
 DR SMART; SM00244; PRB; 1.  
 DR PROSITE; PS01270; BAND 7; 1.  
 KM Transmembrane; Multigene family.  
 FT TRANSMEM 28 48 POTENTIAL.  
 FT DOMAIN 202 238  
 SQ SEQUENCE 281 AA; 30843 MW; DB6648D4682FC9EC CRC64;  
 Query Match 2.7%; Score 7; DB 1; Length 281;

Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 217 TLPLSAF 223  
Db 41 TLPLSAF 47

```
RESULT 10
SC14 CANAL STANDARD; PRT; 301 AA.
ID SC14 CANAL
AC P46250;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SC14 cytosolic factor (Phosphatidylinositol/phosphatidylcholine
transfer protein) (PI/PC TP).
GN SC14.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 64365 / 1001;
RX MEDLINE=97051600; PubMed=8696277;
RA Montecoliva L., Sanchez M., Pla J., Nombela C.;
RT "Cloning of Candida albicans SC14 gene homologue coding for a
RT putative essential function."
RL Yeast 12:1097-1105(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SGY243;
RX MEDLINE=98048477; PubMed=9387231;
RA Riggle P.J., Slobodkin I.V., Brown D.H. Jr., Hanson M.P.,
RA Volkert T.L., Kumamoto C.A.;
RT "Two transcripts, differing at their 3' ends, are produced from the
RT Candida albicans SC14 gene."
RL Microbiology 143:3527-3535(1997).
CC -1- FUNCTION: REQUIRED FOR TRANSPORT OF SECRETORY PROTEINS FROM THE
CC GOLGI COMPLEX. CATALYZES THE TRANSFER OF PHOSPHATIDYLINOSITOL AND
CC PHOSPHATIDYLCHOLINE BETWEEN MEMBRANES IN VITRO (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE GOLGI COMPLEX AS A
CC PERIPHERAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: Contains 1 CRAL-TRIO domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81937; CA54790.1; -
DR EMBL; U61975; AAB4191.1; -
DR HSSP; P24280; 1AUA.
DR InterPro; IPR001251; CRAL_TRIO.
DR InterPro; IPR001071; RetBind/coctrans.
DR Pfam; PF00650; CRAL_TRIO; 1.
DR Pfam; PF03765; CRAL_TRIO_N; 1.
DR PRINTS; PR00180; CRETINALDHP.
DR SMART; SM00516; SEC14; 1.
DR PROSITE; PS50191; CRAL_TRIO; 1.
KM Transport; Protein transport; Golgi stack.
FT DOMAIN 99 272 CRAL-TRIO.
SQ SEQUENCE 301 AA; 34709 MW; 1F194A8E8B5255B CRC64;
```

Query Match 2.7%; Score 7; DB 1; Length 301;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 189 ILGYSYK 195

Db 242 ILGYSYK 248

```
RESULT 11
GPDA_HAEIN STANDARD; PRT; 335 AA.
ID GPDA_HAEIN
AC P43798;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-
dependent glycerol-3-phosphate dehydrogenase).
GN GPDA OR H10605.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kesteven A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McEwen K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ureback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P) (+) =
CC glycerone phosphate + NAD(P)H.
CC -1- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate
CC formation.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32743; AAC22264.1; -
DR PIR; F64080; F64080.
DR TIGR; H10605; -
DR HAMAP; MF_00394; -; 1.
DR InterPro; IPR006168; NAD_Gly3P_dh.
DR InterPro; IPR006109; NAD_Gly3P_dom.
DR Pfam; PF01210; NAD_Gly3P_dh; 1.
DR PRINTS; PR00077; GPDHGRNASE.
DR PRODOM; PD001278; NAD_Gly3P_dom; 1.
DR PROSITE; PS00957; NAD_G3PDH; 1.
KM Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 335 AA; 36394 MW; 4913FBF203B7CFC CRC64;
```

Query Match 2.7%; Score 7; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 IAIAGGI 188  
Db 198 IAIAGGI 204

RESULT 12

```

Y780_METJA          STANDARD;          PRT;          335 AA.
ID Y780_METJA
AC 058130;
RT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M00780.
GN M00780.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;

SEQUENCE FROM N.A.
RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; Pubmed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weisslock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Utechack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurt M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moeze C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67522; AAB98779.1; -.
DR PIR; D64397; D64397.
DR TIGR; M0780; -.
DR Pfam; PF01961; Pfam arch: 1.
KW Hypothetical protein, Transmembrane, Complete proteome.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.
FT TRANSMEM 310 330 POTENTIAL.
SQ SEQUENCE 335 AA; 38138 MM; C6028EDBA1455C08 CRC64;

Query Match 2.7%; Score 7; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 IGAGIIL 190
Db 137 IGAGIIL 143

RESULT 13
ARGC SYNY3
ID ARG_C SYNY3 STANDARD; PRT; 351 AA.
AC P54859;
RT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-
DE acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).
GN ARGC OR SL10080.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; Pubmed=8590279;

```

```

RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64k to 92k of the genome."
RL DNA Res. 2:153-166(1995).
CC -1- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+)
CC + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.
CC -1- PATHWAY: Arginine biosynthesis; third step.
CC -1- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily
CC 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D64004; BAA10557.1; -.
DR PIR; S76613; S76613.
DR HAMAP; MF_00150; -.
DR InterPro; IPR000706; AGPR_act_site.
DR InterPro; IPR000534; Semialdh_dh.
DR Pfam; PF01118; Semialdh_dh; 1.
DR Pfam; PF02774; Semialdh_dh; 1.
DR ProDom; PD003765; AGPR_act_site; 1.
DR PROSITE; PS01224; AGC; 1.
KW Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.
FT ACT SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 351 AA; 38209 MM; C9746F0120E2C8BB CRC64;

Query Match 2.7%; Score 7; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 NMLAEE 19
Db 195 NMLAEE 201

RESULT 14
XYLR_LACPE
ID XYLR_LACPE STANDARD; PRT; 388 AA.
AC P21940;
RT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Xylose repressor.
GN XYLR.
OS Lactobacillus pentosus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1589;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MD353;
RX MEDLINE=92079891; Pubmed=1660563;
RA Lokman B.C., van Santen P., Verdoes J.C., Krusee J., Leer R.J.,
RA Posno M., Pouwels P.H.;
RT "Organization and characterization of three genes involved in
RT D-xylose catabolism in Lactobacillus pentosus."
RL Mol. Gen. Genet. 230:161-169(1991).
CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF XYLOSE-UTILIZING ENZYMES.
CC -1- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



DR	HSSP; P04693; 37A1.	Aminotransf1/2.
DR	InterPro; IPR004839;	Aspartat sub.
DR	InterPro; IPR000796;	NHtrnsf 1.
DR	InterPro; IPR004838;	NHtrnsf 1.
DR	Pfam; PF00155;	aminotran 1-2-1.
DR	PRINTS; PR00799;	TRANSAMINASE.
DR	PROSITE; PS00105;	AA_TRANSFERR_CLASS_1; 1.
DR	Transterase; AminoTransferase;	Pyridoxal phosphate; Complete proteome.
PT	BINDING 248	248
PT	SEQUENCE 398 AA;	43319 MW; 4FE1D5709BAE599F CRC64;
QY	Query Match	2.7%; Score 7; DB 1; Length 398;
QY	Best Local Similarity	100.0%; Pred. No. 31;
QY	Matches 7; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
db	135 NALPARS 141	
db	169 NALPARS 175	

```

RESULT 16
PURA_BRUB ID PURA_BRUB STANDARD; PRT; 402 AA.
AC PS2004; DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adenylosuccinate synthetase [EC 6.3.4.4] (IMP--aspartate ligase)
DE (AdSS) (AMPase) [fragment].
GN PURA.
OC Brucella abortus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxId=235;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2308;
RA Tatum F.M., Steckelberg M.A.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP IDENTIFICATION OF PROBABLE C-TERMINAL FRAMESHIFT.
RA Michoud K.;
RL Unpublished observations (NOV-1999).
CC -1 FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE
CC NUCLEOTIDE BIOSYNTHESIS.
CC -1 CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +
CC adenylosuccinate.
CC -1 PATHWAY: AMP biosynthesis; first committed step.
CC -1 SIMILARITY: Belongs to the adenylosuccinate synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/isb-sib.ch).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; LA3054; AAA75455.1; -.
DR HSPB; PI2283; IADE.
DR HAMAP; MF_00011; -, 1.
DR InterPro; IPR001114; Asucc_synthase.
DR Pfam; PF00709; Adenylsuc_synt; 1.
DR ProDom; PD001188; Asucc_synthase; 1.
DR TRGFAMS; TIGR00184; pura; 1.
DR PROSITE; PS00513; ADENYLOSUCCIN_SYN_2; 1.
DR PROSITE; PS01266; ADENYLOSUCCIN_SYN_1; 1.
KW Purine biosynthesis; Ligase; GTP-binding.
FT NP_BIND 12 18
FT ACT_SITE 140 140
FT ACT_SITE 147 147
FT CONFLICT 393 402 BY SIMILARITY.
FT NON_TER 402 402 XKYVRHIIEEL -> SNNSAISKS (IN REF. 1).

```

SQ SEQUENCE 402 AA; 43919 MW; F443CFAB8D10EE75 CRC64;  
 Query Match 2.7%; Score 7; DB 1; Length 402;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 170 GYVLGIT 176  
 |||||  
 Db 258 GYVLGIT 264  
 RESULT 17  
 PURA BRUME STANDARD; PRT; 429 AA.  
 AC O8Y1T1, Q8FZ20, 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)  
 GN (AdSs) (AMPSase)  
 OS PURA OR BME10351 OR BR1683.  
 OS Brucella melitensis, and  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_Taxid=29459, 29461;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=B.melitensis; STRAIN=16W / ATCC 23456 / Biotype 1;  
 RX MEDLINE=20020109; PubMed=11756588;  
 RA D'Alencio V.G., Kaparakis V., Redkar R.J., Patra G., Mijer C., Los T.,  
 RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Kenik G.,  
 RA Jablonski U., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,  
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
 RA Haselkorn R., Kyrides N., Overbeek R.;  
 RT "The genome sequence of the facultative intracellular pathogen  
 RT Brucella melitensis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=B.suis; STRAIN=1330 / Biovar 1;  
 RX MEDLINE=22247741; PubMed=12271122;  
 RA Paulsen I.T., Seshadri R., Nelson K.E., Bisen J.A., Heidelberg J.F.,  
 RA Read T.D., Dodson R.U., Mayhew L., Brinkac L.M., Beanan M.J.,  
 RA Daugherty S.C., Debay R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,  
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
 RA Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;  
 RT "The Brucella suis genome reveals fundamental similarities between  
 RT animal and plant pathogens and symbionts";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 CC -1- FUNCTION: Plays an important role in the de novo pathway of purine  
 CC nucleotide biosynthesis.  
 CC -1- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +  
 CC adenylosuccinate.  
 CC -1- PATHWAY: AMP biosynthesis; first committed step.  
 CC -1- SIMILARITY: Belongs to the adenylosuccinate synthetase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, AE009477; AAL51532.1; ALT\_INT1.  
 DR EMBL, AE014461; AAN30583.1; -.  
 DR PIR, A13295; A13295.  
 DR TIGR, BR1683; -.  
 DR HAMAP, MF\_000111; -; 1.  
 DR InterPro, IPR001114; Asucc synthetase.  
 DR Pfam, PF00709; Adenylosucc\_synt, 1.

DR ProDom, PD001188; Asucc synthetase; 3.  
 DR TIGRfam, TIGR00184; purA, 1.  
 DR PROSITE, PS01266; ADENYLOSUCIN SYN 1; 1.  
 DR PROSITE, PS00513; ADENYLOSUCIN SYN 2; 1.  
 KW Purine biosynthesis; Ligase; GTP-binding; Complete proteome.  
 FT NP BIND 103 109  
 FT ACT\_SITE 231 231  
 FT ACT\_SITE 238 238  
 FT ACT\_SITE 238 238  
 SQ SEQUENCE 429 AA; 46561 MW; DF90E83E30690PF2 CRC64;  
 Query Match 2.7%; Score 7; DB 1; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 170 GYVLGIT 176  
 |||||  
 Db 259 GYVLGIT 265  
 RESULT 18  
 PURA RHIL0 STANDARD; PRT; 432 AA.  
 AC O98F97;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)  
 GN (AdSs) (AMPSase).  
 OS PURA OR ML3873.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Noduliales; Rhizobiaceae; Mesorhizobium.  
 OX NCBI\_Taxid=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFR30309;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Katsuma T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 CC -1- FUNCTION: Plays an important role in the de novo pathway of purine  
 CC nucleotide biosynthesis.  
 CC -1- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +  
 CC adenylosuccinate.  
 CC -1- PATHWAY: AMP biosynthesis; first committed step.  
 CC -1- SIMILARITY: Belongs to the adenylosuccinate synthetase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, AP003002; BAB50670.1; -.  
 DR HAMAP, MF\_000111; -; 1.  
 DR InterPro, IPR001114; Asucc synthetase.  
 DR Pfam, PF00709; Adenylosucc\_synt, 1.  
 DR ProDom, PD001188; Asucc synthetase; 1.  
 DR TIGRfam, TIGR00184; purA, 1.  
 DR PROSITE, PS01266; ADENYLOSUCIN SYN 1; 1.  
 DR PROSITE, PS00513; ADENYLOSUCIN SYN 2; 1.  
 KW Purine biosynthesis; Ligase; GTP-binding; Complete proteome.  
 FT NP BIND 12 18  
 FT ACT\_SITE 143 143  
 FT ACT\_SITE 150 150  
 FT ACT\_SITE 150 150  
 SQ SEQUENCE 432 AA; 46671 MW; 8726979D4C5E5D07 CRC64;

```

Query Match          2.7%; Score 7; DB 1; Length 432;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      170 GYVLGIT 176
      |||||
Db      262 GYVLGIT 268

RESULT 19
FTSA_PORGI
ID FTSA_PORGI STANDARD; PRT: 456 AA.
AC 007827;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cell division protein ftsa.
GN FTSA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteriia; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RL Akiyama S., Tamura H., Anzai T., Takehana T.,
Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN MAY BE INVOLVED IN ANOMALOUS FILAMENT
CC GROWTH (BY SIMILARITY). MAY BE A COMPONENT OF THE SEPTUM. IT
CC MAY INTERACT WITH FTSS (BY SIMILARITY).
CC -1- SIMILARITY: Belongs to the ftsa/mreB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB004555; BAA20534.1; -
DR InterPro: IPR003494; FtsA.
DR Pfam: PF02491; FtsA; 2.
DR TIGRFAMs: TIGR01174; ftsa; 1.
KW Cell division; Cell shape.
SQ SEQUENCE 456 AA; 5125 MW; 22755F1DF243F6DF CRC64;

Query Match          2.7%; Score 7; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      184 IGAGIIL 190
      |||||
Db      321 IGAGIIL 327

RESULT 20
EMRB_ECOLI
ID EMRB_ECOLI STANDARD; PRT: 512 AA.
AC P27304; P77725;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein B.
GN EMRB OR B2686 OR Z3987 OR BC5548.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562; 83334;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=93028382; PubMed=1409590;
RA Lomovskaya O., Lewis K.;
RT "mr, an Escherichia coli locus for multidrug resistance.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8938-8942(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1234-1238(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Ohshima T., Oyama S., Saito N., Saito Y., Satoh Y., Sivasubraman S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horikuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Pfeif G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamous K.,
RA Apodaca J., Antcharanan T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: TRANSLUCASE THAT CONFERS RESISTANCE TO SUBSTANCES OF
CC HIGH HYDROPHOBICITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLUCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M86657; AAA23725.1; -
DR EMBL: AE000353; AAC75733.1; -
DR EMBL: D90891; BAA16548.1; -
DR EMBL: D90892; BAA16553.1; -
DR EMBL: AE005497; AAG57795.1; -
DR EMBL: AP002562; BAB36971.1; -
DR PTR: D91072; D91072.
DR PTR: G65048; G65048.

```

DR EcGene; EGI1439; emrB.  
 DR InterPro: IPR004638; Efflux\_EmrB.  
 DR InterPro: IPR007114; MFS.  
 DR InterPro: IPR005828; Sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr; 1.  
 DR TIGRPFAMs: TIGR00711; efflux EmrB; 1.  
 KM Antibiotic resistance; Transport; Transmembrane; Inner membrane;  
 Complete proteome.

FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 13 38 POTENTIAL.  
 FT DOMAIN 39 51 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 52 72 POTENTIAL.  
 FT DOMAIN 73 81 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 82 100 POTENTIAL.  
 FT DOMAIN 101 109 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 110 128 POTENTIAL.  
 FT DOMAIN 129 136 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 137 159 POTENTIAL.  
 FT DOMAIN 160 164 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 165 189 POTENTIAL.  
 FT DOMAIN 190 202 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 203 223 POTENTIAL.  
 FT DOMAIN 224 233 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 234 249 POTENTIAL.  
 FT DOMAIN 250 271 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 272 295 POTENTIAL.  
 FT DOMAIN 296 305 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 306 329 POTENTIAL.  
 FT DOMAIN 330 335 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 336 355 POTENTIAL.  
 FT DOMAIN 356 363 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 364 387 POTENTIAL.  
 FT DOMAIN 388 407 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 408 428 POTENTIAL.  
 FT DOMAIN 429 481 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 482 504 POTENTIAL.  
 FT DOMAIN 505 512 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 513 510 POLY-GLY.  
 FT DOMAIN 511 512 G -> A (IN REF. 1).  
 FT CONFLICT 325 325 P -> A (IN REF. 1).  
 FT CONFLICT 501 501  
 SQ SEQUENCE 512 AA; 55624 MW; 0703345688303EFF CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GLASAPV 61  
 |||||  
 Db 307 GLASAPV 313

RESULT 21  
 TY3H HUMAN STANDARD; PRT; 528 AA.  
 ID TY3H HUMAN  
 AC P07101; Q15585; Q15588; Q15589;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).  
 GN TH OR TYH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RP [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).  
 RX MEDLINE=87173064; PubMed=2882428;  
 RA Grima B., Lamouroux A., Boni C., Julien J.-F., Javoy-Agid F.,  
 RA Mallet J.;  
 RT "A single human gene encoding multiple tyrosine hydroxylases with  
 RT different predicted functional characteristics";  
 RL Nature 326:707-711(1987).  
 RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RX MEDLINE=87316931; PubMed=2888085;  
 RA Kobayashi K., Kaneda N., Ichinose H., Kishi F., Nakazawa A.,  
 RA Kurosawa Y., Fujita K., Nagatsu T.;  
 RT "Isolation of a full-length cDNA clone encoding human tyrosine  
 RT hydroxylase type 3";  
 RL Nucleic Acids Res. 15:6733-6733(1987).  
 RN [3]  
 RP SEQUENCE OF 1-94 FROM N.A. (ISOFORM 3).  
 RX MEDLINE=87298614; PubMed=2887169;  
 RA Kaneda N., Kobayashi K., Ichinose H., Kishi F., Nakazawa A.,  
 RA Kurosawa Y., Fujita K., Nagatsu T.;  
 RT "Isolation of a novel cDNA clone for human tyrosine hydroxylase:  
 RT alternative RNA splicing produces four kinds of mRNA from a single  
 RT gene";  
 RL Biochem. Biophys. Res. Commun. 146:971-975(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).  
 RX MEDLINE=89008200; PubMed=2902075;  
 RA Kobayashi K., Kaneda N., Ichinose H., Kishi F., Nakazawa A.,  
 RA Kurosawa Y., Fujita K., Nagatsu T.;  
 RT "Structure of the human tyrosine hydroxylase gene: alternative  
 RT splicing from a single gene accounts for generation of four mRNA  
 RT types";  
 RL J. Biochem. 103:907-912(1988).  
 RN [5]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE=88117543; PubMed=2892893;  
 RA le Bourdellès B., Boularand S., Boni C., Horellou P., Dumas S.,  
 RA Grima B., Mallet J.;  
 RT "Analysis of the 5' region of the human tyrosine hydroxylase gene:  
 RT combinatorial patterns of exon splicing generate multiple regulated  
 RT tyrosine hydroxylase isoforms";  
 RL J. Neurochem. 50:988-991(1988).  
 RN [6]  
 RP SEQUENCE OF 1-30 FROM N.A.  
 RX MEDLINE=88213428; PubMed=2896667;  
 RA Ginns E.I., Rehay M., Martin B.M., Weller M., O'Malley K.L.,  
 RA Lamarc M.E., McAllister C.G., Paul S.M.;  
 RT "Expression of human tyrosine hydroxylase cDNA in invertebrate cells  
 RT using a baculovirus vector";  
 RL J. Biol. Chem. 263:7406-7410(1988).  
 RN [7]  
 RP VARIANT SEGAWA SYNDROME LYS-412.  
 RX MEDLINE=95113436; PubMed=7814018;  
 RA Luedecke B., Dworniczak B., Bartholome K.;  
 RT "A point mutation in the tyrosine hydroxylase gene associated with  
 RT Segawa's syndrome";  
 RL Hum. Genet. 95:123-125(1995).  
 RN [8]  
 RP VARIANT MET-112.  
 RX MEDLINE=95309895; PubMed=7789962;  
 RA Luedecke B., Bartholome K.;  
 RT "Frequent sequence variant in the human tyrosine hydroxylase gene";  
 RL Hum. Genet. 95:716-716(1995).  
 RN [9]  
 RP CHARACTERIZATION OF VARIANT SEGAWA SYNDROME LYS-412.  
 RX MEDLINE=96133297; PubMed=8528210;  
 RA Knappskog P.M., Flatmark T., Mallet J., Luedecke B., Bartholome K.;  
 RT "Recessively inherited L-DOPA-responsive dystonia caused by a point  
 RT mutation (Q381K) in the tyrosine hydroxylase gene";  
 RL Hum. Mol. Genet. 4:1209-1212(1995).  
 RN [10]  
 RP CHARACTERIZATION OF VARIANT SEGAWA SYNDROME PRO-236.  
 RX MEDLINE=96414313; PubMed=8817341;  
 RA Luedecke B., Knappskog P.M., Clayton P.T., Surtrees R.A.H.,  
 RA Clelland J.D., Heales S.J.R., Brand M.P., Bartholome K., Flatmark T.;  
 RT "Recessively inherited L-DOPA-responsive parkinsonism in infancy  
 RT caused by a point mutation (L205P) in the tyrosine hydroxylase gene";  
 RL Hum. Mol. Genet. 5:1023-1028(1996).  
 RN [11]  
 RP VARIANT SEGAWA SYNDROME PRO-236, AND VARIANT MET-112.  
 RX MEDLINE=98273932; PubMed=9613851;

RA Kunugi H., Kawada Y., Hattori M., Ueki A., Otsuka M., Nanko S.;  
 RT "Association study of structural mutations of the tyrosine hydroxylase  
 RT gene with schizophrenia and Parkinson's disease.";  
 RL Am. J. Med. Genet. 81:131-133(1998).  
 RN [12]  
 RP VARIANT MET-499.  
 RX MEDLINE=98425595; PubMed=9754624;  
 RA Ishiguro H., Arinami T., Saito T., Akazawa S., Enomoto M.,  
 RA Ichisato H., Fujishiro H., Tada K., Akimoto Y., Mitune H.,  
 RA Shozuka S., Hamaguchi H., Toru M., Shibuya H.;  
 RT "Systematic search for variations in the tyrosine hydroxylase gene and  
 RT their associations with schizophrenia, affective disorders, and  
 RT alcoholism.";  
 RL Am. J. Med. Genet. 81:388-396(1998).  
 RN [13]  
 RP VARIANT SEGAWA SYNDROME HIS-233.  
 RX MEDLINE=98366985; PubMed=9703425;  
 RA van den Heuvel L.P.W.J., Luiten B., Smeitink J.A.M.,  
 RA de Rijk-van Andel J.F., Hyland K., Steenbergen-Spanjers G.C.H.,  
 RA Tanssen R.J.T., Wevers R.A.;  
 RT "A common point mutation in the tyrosine hydroxylase gene in autosomal  
 RT recessive L-DOPA-responsive dystonia in the Dutch population.";  
 RL Hum. Genet. 102:644-646(1998).  
 RN [14]  
 RP VARIANT MET-112.  
 RX MEDLINE=99318093; PubMed=10391209;  
 RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.O.,  
 RA Lander E.S.;  
 RT "Characterization of single-nucleotide polymorphisms in coding regions  
 RT of human genes.";  
 RL Nat. Genet. 22:231-238(1999).  
 RN [15]  
 RP ERRATUM.  
 RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.O.,  
 RA Lander E.S.;  
 RL Nat. Genet. 23:373-373(1999).  
 RN [16]  
 RP VARIANT SEGAWA SYNDROME PRO-276; MET-314; HIS-337 AND MET-494.  
 RX MEDLINE=21141790; PubMed=11246459;  
 RA Swaans R.J.M., Rondot P., Renier W.O., Van Den Heuvel L.P.W.J.,  
 RA Steenbergen-Spanjers G.C.H., Wevers R.A.;  
 RT "Four novel mutations in the tyrosine hydroxylase gene in patients  
 RT with infantile parkinsonism.";  
 RL Ann. Hum. Genet. 64:25-31(2000).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC  
 CC NEURONES.  
 CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-  
 CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.  
 CC -1- COFACTOR: FERROUS ION.  
 CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE  
 CC CATALYTIC ACTIVITY.  
 CC -1- PATHWAY: Catecholamine biosynthesis; first step.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Name=3;  
 CC IsoId=P07101-1; Sequence=Displayed;  
 CC Name=1;  
 CC IsoId=P07101-2; Sequence=VSP\_000543;  
 CC Name=2;  
 CC IsoId=P07101-3; Sequence=VSP\_000544;  
 CC Name=4;  
 CC IsoId=P07101-4; Sequence=VSP\_000541, VSP\_000542;  
 CC -1- TISSUE SPECIFICITY: Mainly expressed in the brain and adrenal  
 CC glands.  
 CC -1- DISEASE: Defects in TH are the cause of autosomal recessive Segawa  
 CC syndrome [MIM:605407]; also known as dopa-responsive dystonia.  
 CC Typically, it begins in childhood or adolescence with progressive  
 CC difficulty in walking and, in some cases, spasticity. Some cases  
 CC present with parkinsonian symptoms in infancy and are referred to

CC as autosomal recessive infantile parkinsonism.  
 CC -1- SIMILARITY: Belongs to the bioprotein-dependent aromatic amino acid  
 CC hydroxylase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y00414; CAA68472.1; ALT\_SEQ.  
 CC EMBL: M17589; AAA61179.1; -.  
 CC EMBL: X05290; CAA28908.1; -.  
 CC EMBL: M24791; AAA61173.1; -.  
 CC EMBL: M24787; AAA61173.1; JOINED.  
 CC EMBL: M24789; AAA61173.1; JOINED.  
 CC EMBL: M24791; AAA61170.1; -.  
 CC EMBL: M24787; AAA61170.1; JOINED.  
 CC EMBL: M20911; AAA61167.1; -.  
 CC PIR: A30002; WHTU04.  
 CC HSP: P04177; 1T0H.  
 CC DR Genew: HGNC:11782; TH.  
 CC DR MIM: 191290; -.  
 CC DR MIM: 605407; -.  
 CC DR GO: GO:0004511; P-tyrosine 3-monooxygenase activity; TAS.  
 CC DR GO: GO:0007345; P-embryogenesis and morphogenesis; TAS.  
 CC DR InterPro: IPR001273; Aaa hydroxylase.  
 CC DR InterPro: IPR005962; Tyr 3 monox.  
 CC DR Pfam: PF00351; bioprotein\_H\_1.  
 CC DR PRINTS: PR00372; FYWHYDRKLASE.  
 CC DR ProDom: PD002559; Aaa hydroxylase; 1.  
 CC DR TIGRFAMs: TIGR01269; Tyr\_3\_monoox; 1.  
 CC -----  
 CC Query Match 2.7%; Score 7; DB 1; Length 528;  
 CC Best Local Similarity 100.0%; Pred. No. 39;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC Qy 142 EAAAVOP 148  
 CC Db 446 EAAAVOP 452  
 CC |||||  
 CC  
 CC RESULT 22  
 CC ID UREL PROMI STANDARD; PRT; 567 AA.  
 CC AC P17086;  
 CC DT 01-AUG-1990 (Rel. 15, Created)  
 CC DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Urease alpha subunit (EC 3.5.1.5) (Urea amidohydrolase).  
 CC GN UREAC.  
 CC OS Proteus mirabilis.  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC OC Enterobacteriaceae; Proteus.  
 CC OX NCBI\_TaxID=584;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=H14320;  
 CC RX MEDLINE=90078080; PubMed=2687233;  
 CC RA Jones B.D., Modley H.U.T.;  
 CC RT "Protein mirabilis urease: nucleotide sequence determination and  
 CC comparison with jack bean urease.";  
 CC RL J. Bacteriol. 171:6414-6422(1989).  
 CC CC -1- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).  
 CC -1- COFACTOR: Binds 2 nickel ions per subunit (By similarity).  
 CC -1- SUBUNIT: (ALPHA, BETA, GAMMA) (3) (By similarity).  
 CC -1- PTM: Lys-217 is carbamylated. The carbamoyl group provides the  
 CC ligands for the two nickel ions (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE UREASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31834; AAA25669.1; -
CC PIR: D43719; DA3719.
CC HSSP: P18314; 1PWE.
CC MEROPS: M38.UNM; -
CC InterPro: IPR006680; Amidohydro_1.
CC InterPro: IPR005847; Urease.
CC InterPro: IPR005848; Ureasea.
CC Pfam: PF01979; Amidohydro_1; 1.
CC Pfam: PF00449; Urease; 1.
CC PROSITE: PS00145; UREASE_2; 1.
CC PROSITE: PS01120; UREASE_1; 1.
CC Hydroxylase; Metal-binding; Nickel.
CC METAL 134 134 NICKEL 2 (BY SIMILARITY).
CC METAL 136 136 NICKEL 2 (BY SIMILARITY).
CC METAL 217 217 NICKEL 1 AND 2 (BY SIMILARITY).
CC METAL 246 246 NICKEL 1 (BY SIMILARITY).
CC METAL 272 272 NICKEL 1 (BY SIMILARITY).
CC METAL 360 360 NICKEL 2 (BY SIMILARITY).
CC ACT SITE 320 320 BY SIMILARITY.
CC SEQUENCE 567 AA; 61012 MW; 67794F5D95A8E1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 567;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 EAGVPEK 93
DB 498 EAGVPEK 504

RESULT 23
ID ILVB_LACLA STANDARD; PRT; 575 AA.
AC Q02137; Q9CG84;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Acetolactate synthase large subunit (EC 2.2.1.6) (AHAS)
DE (Acetylhydroxy-acid synthase large subunit) (ALS).
GN ILVB OR ILI224.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis)
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 2118; PubMed=1400210;
RX MEDLINE=93015710; PubMed=1400210;
RA Godon J.-J., Choplin M.-C., Ehrlich S.D.;
RT "Branched-chain amino acid biosynthesis genes in Lactococcus lactis
RT subsp. lactis";
RL J. Bacteriol. 174:6580-6589(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ILI403.
RX MEDLINE=21235186; PubMed=11337471;
RA Bojotin A., Wincker P., Mauger S., Jallion O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis ILI403";
RL Genome Res. 11:731-753(2001).
RN [3]
CC -1- CATALYTIC ACTIVITY: 2 pyruvate = 2-acetolactate + CO(2).
CC -1- COFACTOR: Binds 1 thiamine pyrophosphate and 1 magnesium ion per
CC subunit (by similarity).
CC -1- PATHWAY: Valine and isoleucine biosynthesis; first step.
CC -1- SUBUNIT: DIMER OF LARGE AND SMALL CHAINS.
CC -1- SIMILARITY: BELONGS TO THE TTP ENZYMES FAMILY.

```

```

CC -1- CAUTION: REF 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A STOP
CC CODON IN POSITION 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U92974; AAB81919.1; -
CC EMBL: AE006354; AAK0322.1; ALT_INIT.
CC PIR: S35138; S35138.
CC HSSP: P06169; 1PVD.
CC InterPro: IPR004407; Acclac_19.
CC InterPro: IPR000399; Pyruvate_decarb.
CC Pfam: PF00205; TTP_enzymes; 1.
CC Pfam: PF02775; TTP_enzymes_C; 1.
CC Pfam: PF02776; TTP_enzymes_N; 1.
CC DR TIGRfams; TIGR00118; acclac_19; 1.
CC PROSITE: PS00187; TTP_ENZYMES; 1.
CC Transferrase; Branched-chain amino acid biosynthesis; Flavoprotein;
CC Magnesium; Thiamine pyrophosphate; Complete proteome.
CC ACT SITE 57 57 BY SIMILARITY.
CC CONFLICT 111 112 PL -> RQ (IN REF. 1).
CC CONFLICT 298 298 V -> F (IN REF. 1).
CC CONFLICT 353 354 TK -> IE (IN REF. 1).
CC CONFLICT 557 557 S -> N (IN REF. 1).
CC CONFLICT 568 568 E -> K (IN REF. 1).
CC CONFLICT 572 572 V -> I (IN REF. 1).
CC SEQUENCE 575 AA; 62849 MW; 2F7168C50409035F CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 575;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LIAEAYG 21
DB 504 LIAEAYG 510

RESULT 24
ID PU91_YEAST STANDARD; PRT; 591 AA.
AC P54113;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bifunctional purine biosynthesis protein ADE16 [includes:
DE Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)
DE (IMP synthetase) (ATIC)].
GN ADE16 OR YLR028C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Tibbetts A.S.;
RT Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansgorge W.,
RA Benes V., Brueckner M., Delius H., Dubois B., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miesga T., Moestl D.,
RA Mueller-Aer S., Nentwich U., Obermaier B., Pitravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M.,

```

```

RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Uegles K., Lye G., Moule S., Odell C., Pearson D., Rajadream M.A.,
RA Rice P., Skellern J., Walsh S., Whitehead S., Barrall B.G.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";
RL Nature 387:90-93(1997).
RN
RP SEQUENCE OF 140-150 AND 389-400.
RC STRAIN=5288C;
RX MEDLINE=95203288; PubMed=7895733;
RA Garrels J.I., Fletcher B., Kobayashi R., Latter G.I., Schwender B.,
RA Volpe T., Warner U.R., McLaughlin C.S.;
RT "Protein identifications for a Saccharomyces cerevisiae protein
RL database.";
RL Electrophoresis 15:1466-1486(1994).
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-
CC phospho-D-ribose)imidazole-4-carboxamide = tetrahydrofolate + 5-
CC formamido-1-(5-phospho-D-ribose)imidazole-4-carboxamide.
CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-phospho-D-
CC ribosyl)imidazole-4-carboxamide.
CC -1- PATHWAY: De novo purine biosynthesis; ninth step.
CC -1- PATHWAY: De novo purine biosynthesis; tenth step.
CC -1- SUBUNIT: HOMODIMER (POSSIBLE).
CC -1- DOMAIN: THE IMP CYCLOHYDROLASE ACTIVITY RESIDES IN THE N-TERMINAL
CC REGION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PURH FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49273; CAA89269.1; -.
DR PIR; S54489; S54489.
DR SCOP; S0004727; ADE17.
DR GO; GO:0005829; C:Cytosol; IDA.
DR GO; GO:0003937; F:IMP cyclohydrolase activity; IDA.
DR GO; GO:0004443; F:phosphoribosylaminoimidazole-carboxamide fo. .; IDA.
DR GO; GO:0006189; P:de novo IMP biosynthesis; IDA.
DR InterPro; IPR002695; AICARFT IMCChas.
DR InterPro; IPR004362; MGS like.
DR Pfam; PF01808; AICARFT_IMCChas; 1.
DR Pfam; PF02142; MGS; 1.
DR ProDom; PD004666; AICARFT_IMCChas; 1.
DR TRIPFams; TIGR00355; PuriH; 1.
KM Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme.
PT CONFLICT 389 389 R -> A (IN REF. 2).
SQ SEQUENCE 592 AA; 65263 MW; 8ABA1761B512242 CRC64;
Query Match 2.7%; Score 7; DB 1; Length 592;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 198 KDLKEQH 204
DB 87 KDLKEQH 93

```

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxId=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia; TISSUE=leaf;  
RX MEDLINE=93192524; PubMed=8448363;  
RA Nordlin K., Vahala T., Palva E.T.;  
RT "differential expression of two related, low-temperature-induced  
RL genes in Arabidopsis thaliana (L.) Heynh.";  
RN Plant Mol. Biol. 21:641-653(1993).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=94143472; PubMed=8310052;  
RA Yamaguchi-Shinozaki K., Shinozaki K.;  
RT "Arabidopsis DNA encoding two desiccation-responsive rd29 genes";  
RL Plant Physiol. 101:1119-1120(1993).  
[3]  
RP SEQUENCE FROM N.A., AND INDUCTION.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=93173109; PubMed=8437577;  
RA Yamaguchi-Shinozaki K., Shinozaki K.;  
RT "Characterization of the expression of a desiccation-responsive rd29  
RL gene of Arabidopsis thaliana and analysis of its promoter in  
RN transgenic plants.";  
RN Mol. Gen. Genet. 236:331-340(1993).  
[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20181125; PubMed=10718197;  
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
RL features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
RN clones.";  
RL DNA Res. 7:31-63(2000).  
[5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the  
RL SSP consortium (Sals/Stanford/PGSC).";  
RN Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.  
[6]  
RP SEQUENCE OF 1-121 FROM N.A.  
RC STRAIN=cv. Columbia; TISSUE=Dry seed;  
RA Raynal M., Gelliet F., Laidie M., Meyer Y., Cooke R., Delzeny M.;  
RL Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases  
CC -I- INDUCTION: BY LOW TEMPERATURE, AND MOSTLY BY WATER STRESS OR  
CC ABSISIC ACID (ABA).  
CC -I- SIMILARITY: BELONGS TO THE LTI78/LTI65 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X67670; CAA47902.1; -  
DR EMBL; D13044; BAA02375.1; -  
DR EMBL; AB019226; BAB10527.1; -  
DR EMBL; AY081282; AAL91171.1; -  
DR EMBL; Z34014; CAA83975.1; -  
DR PIR; S30153; S30153.  
DR Multigene family; Repeat.  
FT DOMAIN 24 27 POLY-GLU.  
FT DOMAIN 64 67 POLY-ASP.  
FT DOMAIN 404 511 5 X 5 AA REPEATS OF [IV]-[AMS]-[EST]-

FT	REPEAT	404	408	K-L.
FT	REPEAT	442	446	1.
FT	REPEAT	460	464	2.
FT	REPEAT	490	494	3.
FT	REPEAT	507	511	4.
FT	DOMAIN	514	517	5.
FT	CONFLICT	24	43	POLY-GLY.
FT	CONFLICT	80	87	MISSING (IN REF. 2 AND 3).
FT	CONFLICT	82	83	PYRSSAV -> T (IN REF. 2 AND 3).
FT	CONFLICT	260	261	YE -> FD (IN REF. 6).
FT	CONFLICT	414	414	MISSING (IN REF. 1, 2, 3 AND 4).
FT	CONFLICT	423	423	N -> L (IN REF. 2 AND 3).
FT	CONFLICT	594	619	V -> L (IN REF. 2 AND 3).
FT	CONFLICT			TMGPSDSGSGSLTGGSGGGKGVQDSEGN -> KTPSSLYCT (IN REF. 1).
FT	CONFLICT	594	619	TMGPSDSGSGSLTGGSGGGKGVQDSEGN -> KISLVAVTRN VKILMCNVF (IN REF. 2 AND 3).
SQ	SEQUENCE	619 AA;	65971 MW;	F2D2DF9C5990A00A CRC64.
Query Match				
Best Local Similarity		2.7%;	Score 7;	DB 1;
Matches		7;	Conservative	0;
			Mismatches	0;
			Indels	0;
			Gaps	0;
OY	86 GEAGVPE 92			
Db	261 GEAGVPE 267			
RESULT 27				
UBA3_HUMAN	STANDARD;	PRT;	661 AA.	
ID UBA3_HUMAN	PS7075;			
AC	16-OCT-2001 (Rel. 40,	Created)		
DT	16-OCT-2001 (Rel. 40,	Last sequence update)		
DT	28-FEB-2003 (Rel. 41,	Last annotation update)		
DE	UBASH3A protein.			
GN	UBASH3A.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OK	NCBI_TaxID=9606;			
RN	[1]			
RA	SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).			
RC	TISSUE=Placenta;			
RX	MEDLINE=21177060; PubMed=11281453;			
RA	Wattenhofer M., Shibuya K., Kudon J., Lyle R., Michaud J., Rossier C., Kawasaki K., Asakawa S., Minoshima S., Berry A., Bonne-Tamir B., Shmizl N., Antonarakis S.E., Scott H.S.;			
RT	"Isolation and characterization of the UBASH3A gene on 21q22.3 encoding a potential nuclear protein with a novel combination of domains";			
RL	Hum. Genet. 108:140-147(2001).			
CC	- SUBCELLULAR LOCATION: Nuclear (Potential).			
CC	- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Name=Long;			
CC	IsoId=P57075-1; Sequence=Displayed;			
CC	Name=Short;			
CC	IsoId=P57075-2; Sequence=VSP_006703;			
CC	- TISSUE SPECIFICITY: Highest expression of UBASH3A in tissues belonging to the immune system, including spleen, peripheral blood leukocytes, and bone marrow.			
CC	- SIMILARITY: Contains 1 SH3 domain.			
CC	- SIMILARITY: Contains 1 UBA domain.			
CC	- SIMILARITY: STRONG, TO DROSOPHILA CG13604.			
CC	-			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcement/ or send an email to license@ebi.ac.uk).			



```
CC -----
DR EMBL: AJ277750; CAB91543.1; -.
DR Genew: HGNC:12462; UBASH3A.
DR MIM: 605736; -.
DR InterPro: IPR001345; PG/BPM_mutase.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR000449; UBA_domain.
DR Pfam: PF00300; PGAM; 1.
DR Pfam: PF00018; SH3; 1.
DR Pfam: PF00627; UBA; 1.
DR Prodom: PD000066; SH3; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
DR PROSITE: PS50030; UBA; 1.
DR Nuclein: protein; SH3 domain; Alternative splicing.
DR DOMAIN 15 60 UBA.
DR DOMAIN 276 340 SH3.
DR VARSPIC 185 223 GTSVSRFVPSQVPGHGNRLSLNLTASVSHYILQKY
FT -> D (in isoform Short).
FT /FTid=VSP_006703.
SQ SEQUENCE 661 AA; 74122 MW; 60DA2E0B8CE4ABFC CRC64;

Query Match 2.7%; Score 7; DB 1; Length 661;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 VIGISOR 155
Db 318 VIGISOR 324

RESULT 28
LEW3 SHEEP STANDARD; PRT; 769 AA.
ID LEW3 SHEEP STANDARD; PRT; 769 AA.
AC P98109;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P-selectin precursor (granule membrane protein 140) (GMP-140) (PADGEM)
DE (CD62P) (leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN SELP.
OS Ovis aries (sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_Taxid=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Burns S.A.; Neufeld E.J.; Donady J.J.;
RL Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 8 Sushi (SCR) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34270; AAB59261.1; -.
DR HSSP: P16109; IPRB.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR001304; Lectin_C.
```

```
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00008; EGF_1.
DR Pfam: PF00059; Lectin_C; 1.
DR Pfam: PF00084; sushi; 8.
DR SMART: SM00032; CCP; 8.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS0186; EGF_2; 1.
DR PROSITE: PS00615; C-TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LLECTIN_2; 1.
DR Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sushi; Repeat.
FT SIGNAL 1 32
FT CHAIN 33 769
FT DOMAIN 33 717
FT TRANSMEM 718 734
FT DOMAIN 735 769
FT DOMAIN 58 158
FT DOMAIN 159 195
FT DOMAIN 199 258
FT DOMAIN 261 320
FT DOMAIN 323 382
FT DOMAIN 385 444
FT DOMAIN 447 505
FT DOMAIN 509 568
FT DOMAIN 580 639
FT DOMAIN 642 701
FT DISULFID 60 158
FT DISULFID 131 150
FT DISULFID 163 174
FT DISULFID 168 183
FT DISULFID 185 194
FT DISULFID 200 244
FT DISULFID 230 257
FT DISULFID 262 306
FT DISULFID 292 319
FT DISULFID 324 368
FT DISULFID 354 381
FT DISULFID 386 430
FT DISULFID 416 443
FT DISULFID 448 492
FT DISULFID 478 505
FT DISULFID 510 554
FT DISULFID 540 567
FT DISULFID 581 625
FT DISULFID 611 638
FT DISULFID 643 687
FT DISULFID 673 700
FT CARBOHYD 54
FT CARBOHYD 80
FT CARBOHYD 180
FT CARBOHYD 212
FT CARBOHYD 219
FT CARBOHYD 347
FT CARBOHYD 398
FT CARBOHYD 604
FT CARBOHYD 655
FT CARBOHYD 662
FT CARBOHYD 680
FT SITE 757 760
FT VARIANT 566 566
FT VARIANT 579 579
SQ SEQUENCE 769 AA; 84317 MW; 23E4575D60FAB15 CRC64;

Query Match 2.7%; Score 7; DB 1; Length 769;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 PAPGLRC 46
Db 575 PAPGLRC 581
```

```

RESULT 29
CUL3 SCHPO STANDARD; PRT; 785 AA.
ID CUL3 SCHPO STANDARD; PRT; 785 AA.
AC 009760; 074185;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cullin 3 homolog (Cul-3).
GN PCU3 OR SPAC2446.03.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,
RA Moore P., Moule S., Mungall K., Murphy L., Niblett D., Odel C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkert A., Aert R., Robben J., Grymmer B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabl C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Mambert R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleaux V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Cerrutski L., Lowe T., McCombie W.R., Armstrong J., Forsburg S.L.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe.";
RT Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 136-543 FROM N.A.
RA Komnani K., Toda T.;
RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: Belongs to the cullin family.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 513
CC ONWARD AND IS SHORTER (421 AA) DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z54142; CA90847.1; -
DR EMBL; AB017028; BAA32519.1; ALT_FRAME.
DR PIR; S62405; S62405.
DR PIR; S38359; S38359.
DR GenDB SPombe; SPAC2446.03; -
DR InterPro; IPR001373; Cullin.
DR Pfam; PF00888; Cullin; 1.
DR SMART; SM00182; CULLIN; 1.
DR PROSITE; PS01256; CULLIN_1; 1.
DR PROSITE; PS50069; CULLIN_2; 2.
FT CONFLICT 476 476 R -> RYALIVETVNTFR (IN REF. 2).
SQ SEQUENCE 785 AA; 90511 MW; 0830F7A857DF6CE7 CRC64;

```

```

Query Match 2.7%; Score 7; DB 1; Length 785;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 ERTSOAL 109
Db 353 ERTSOAL 359

RESULT 30
ROR1 HUMAN STANDARD; PRT; 937 AA.
ID ROR1 HUMAN STANDARD; PRT; 937 AA.
AC 001973; Q92776;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tyrosine-protein kinase transmembrane receptor ROR1 precursor
DE (EC 2.7.1.112) (Neurotrophic tyrosine kinase, receptor-related 1).
GN ROR1 OR NTRK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RX MEDLINE=93100347; PubMed=134494;
RA Maslakowski P., Carroll R.D.;
RT "A novel family of cell surface receptors with tyrosine kinase-like
RT domain.";
RL J. Biol. Chem. 267:26181-26190(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RX MEDLINE=97030043; PubMed=887595;
RA Reddy U.R., Phatak S., Pleasure D.;
RT "Human neural tissues express a truncated Ror1 receptor tyrosine
RT kinase, lacking both extracellular and transmembrane domains.";
RL Oncogene 13:1555-1559(1996).
CC -1- FUNCTION: TYROSINE-PROTEIN KINASE RECEPTOR WHOSE ROLE IS NOT YET
CC CLEAR.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q01973-1; Sequence=Displayed;
CC Name=Short; Synonyms=T-ROR1;
CC IsoId=Q01973-2; Sequence=VSP_005008;
CC -1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN HUMAN HEART, LUNG, AND
CC KIDNEY, BUT WEAKLY IN THE CNS. THE SHORT ISOFORM IS STRONGLY
CC EXPRESSED IN FETAL AND ADULT CNS AND IN A VARIETY OF HUMAN
CC CANCERS, INCLUDING THOSE ORIGINATING FROM CNS OR PNS
CC NEUROCTODEM.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT HIGH LEVELS DURING EARLY
CC EMBRYONIC DEVELOPMENT. THE EXPRESSION LEVELS DROP STRONGLY AROUND
CC DAY 16 AND THERE ARE ONLY VERY LOW LEVELS IN ADULT TISSUES.
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. ROR
CC SUBFAMILY.
CC -1- SIMILARITY: Contains 1 frizzled (FZ) domain.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC -1- SIMILARITY: Contains 1 kringle domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M97675; AAA60275.1; -
DR EMBL; U38894; AAC50714.1; -
DR PIR; A45082; A45082.

```

DR HSSP; P00747; 1CEA.  
 DR Genew; HGNC:10256; ROR1.  
 DR MIM; 602336; -.  
 DR GO; GO:0005737; C:cytoplasm; TAS.  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0004714; P:transmembrane receptor protein tyrosine kin. . . ; TAS.  
 DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . . ; TAS.  
 DR InterPro; IPR000024; Fz\_domain.  
 DR InterPro; IPR007110; Ig\_Like.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR000001; Kringle.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR01245; Tyr\_kinase.  
 DR Pfam; PF01392; Fz; 1.  
 DR Pfam; PF00047; Ig; 1.  
 DR Pfam; PF00051; kringle; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000395; Kringle; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00408; IGC2; 1.  
 DR SMART; SM00130; KR; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS50038; FZ; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR PROSITE; PS00021; KRINGLE; 1; 1.  
 DR PROSITE; PS50070; KRINGLE\_2; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR Transferrase; Tyrosine-protein kinase; ATP-binding; Receptor;  
 KW Transmembrane; Signal; Glycoprotein; Kringle; Phosphorylation;  
 KM Immunoglobulin domain; Alternative splicing.  
 FT CHAIN 1 29  
 FT SIGNAL 30 937  
 FT DOMAIN 30 406  
 FT TRANSFER 407 427  
 FT DOMAIN 428 937  
 FT DOMAIN 42 147  
 FT DOMAIN 165 299  
 FT DOMAIN 312 391  
 FT DOMAIN 473 746  
 FT DOMAIN 753 782  
 FT DOMAIN 784 851  
 FT DOMAIN 853 876  
 FT NP BIND 479 487  
 FT BINDING 506 506  
 FT ACT\_SITE 615 615  
 FT MOD\_RES 645 645  
 FT DISULFID 79 131  
 FT CARBOHYD 47 47  
 FT CARBOHYD 66 66  
 FT CARBOHYD 184 184  
 FT CARBOHYD 315 315  
 FT VARSPIC 1 549  
 SQ SEQUENCE 937 AA; 104312 MW; 0D0694DBF29F4773 CRC64;

Query Match 2.78; Score 7; DB 1; Length 937;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 HSYCRNP 73  
 Db 359 HSYCRNP 365

RESULT 31  
 ROR1\_MOUSE  
 ID ROR1\_MOUSE STANDARD; PRT; 937 AA.

AC Q92139;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Tyrosine-protein kinase transmembrane receptor ROR1 precursor  
 DE (EC 2.7.1.112) (Neurotrophic tyrosine kinase, receptor-related 1)  
 DE (MORF1).  
 OS ROR1 OR NTRK1.  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9248426; PubMed=10231392;  
 RA Oishi I., Takeuchi S., Hashimoto R., Nagabukuro A., Ueda T., Liu Z.J.,  
 RA Hattori T., Akira S., Matsuda Y., Yamamura H., Otani H., Minami Y.,  
 RT "spatio-temporally regulated expression of receptor tyrosine kinases,  
 RT mRor1, mRor2, during mouse development: implications in development  
 RT and function of the nervous system.";  
 RL Genes Cells 4:41-56 (1999).  
 CC -1- FUNCTION: TYROSINE-PROTEIN KINASE RECEPTOR WHOSE ROLE IS NOT YET  
 CC CLEAR.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. FOR  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: Contains 1 frizzled (FZ) domain.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -1- SIMILARITY: Contains 1 kringle domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB010383; BAA75480.1; -.  
 DR HSSP; P00747; 1CEA.  
 DR MGD; MG11347520; Ror1.  
 DR InterPro; IPR000024; Fz\_domain.  
 DR InterPro; IPR007110; Ig\_Like.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR000001; Kringle.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR01245; Tyr\_kinase.  
 DR Pfam; PF01392; Fz; 1.  
 DR Pfam; PF00047; Ig; 1.  
 DR Pfam; PF00051; kringle; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000395; Kringle; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00408; IGC2; 1.  
 DR SMART; SM00130; KR; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS50038; FZ; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR PROSITE; PS00021; KRINGLE; 1; 1.  
 DR PROSITE; PS50070; KRINGLE\_2; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_TYR; 1.  
 DR Transferrase; Tyrosine-protein kinase; ATP-binding; Receptor;  
 KW Transmembrane; Signal; Glycoprotein; Kringle; Phosphorylation;  
 KM Immunoglobulin domain.  
 FT CHAIN 1 29  
 FT SIGNAL 30 937  
 FT TYROSINE-PROTEIN KINASE TRANSMEMBRANE

```

FT DOMAIN 30 406 RECEPTOR ROR1.
FT TRANSMEM 407 427 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 428 937 POTENTIAL.
FT DOMAIN 428 937 CYTOSOLASMIC (POTENTIAL).
FT DOMAIN 165 141 IG-LIKE C2-TYPE.
FT DOMAIN 165 299 F2.
FT DOMAIN 312 391 KRINGLE.
FT DOMAIN 473 746 PROTEIN KINASE.
FT DOMAIN 753 782 SER/THR-RICH.
FT DOMAIN 784 851 PRO-RICH.
FT DOMAIN 853 876 SER/THR-RICH.
FT NE-BIND 479 487 ATP (BY SIMILARITY).
FT BINDING 506 506 ATP (BY SIMILARITY).
FT ACT SITE 615 615 BY SIMILARITY.
FT MOD RES 645 645 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT DISULFID 79 131 BY SIMILARITY.
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 937 AA; 104156 MW; D728733E67D1782C CRC64;

Query Match 2.7%; Score 7; DB 1; Length 937;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 HSYCRNP 73
Db 359 HSYCRNP 365

RESULT 32
ADD DROME STANDARD; PRT; 1034 AA.
AC P54362; O16015; O45031; Q9VY96;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable delta-adaptin (garnet protein) (Delta adaptin subunit of AP-3).
GN 3).
OS G OR CG10986/CG11197.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCB1_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=97447555; PubMed=9103295;
RA Ooi C.E., Moreira J.E., Dell'Angelica E.C., Poy G., Wassarman D.A.,
RA Bonifacio J.S.;
RA "Altered expression of a novel adaptin leads to defective pigment
RT granule biogenesis in the Drosophila eye color mutant garnet.";
RL EMO J. 16:4508-4518(1997).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=20123182; PubMed=10659786;
RA Lloyd V.K., Sinclair D.A., Wennerg R., Warner T.S., Honda B.M.,
RA Grigliatti T.A.;
RT "A genetic and molecular characterization of the garnet gene of
RL Drosophila melanogaster.";
RL Genome 42:1183-1193(1999).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=20123182; PubMed=10659786;
RA Lloyd V.K.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN (4)
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;

```

```

RA Adams M.D., Cejniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scheer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abiri J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Broxstein P., Brothier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Dayenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doop L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mikhlin N.V., Mobarry C., Morris J., Mostrel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Shen H.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RL "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195(2000).
RN (5)
RP REVISIONS.
RC STRAIN=Berkley;
RX MEDLINE=22426069; PubMed=12537572;
RA MEDLINE=22426069; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Cejniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RL systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC -!- FUNCTION: Part of the AP-3 complex, an adaptor-related complex
CC which is not clathrin-associated. The complex is associated with
CC the Golgi region as well as more peripheral structures. It
CC facilitates the budding of vesicles from the Golgi membrane and
CC may be directly involved in trafficking to lysosomes (By
CC similarity).
CC -!- FUNCTION: May be a coat protein involved in the formation of
CC specialized structures like pigment granules.
CC -!- SUBUNIT: Assembly protein complex 3 (AP-3) is a heterotrimer
CC composed of two large chains (delta and beta3), a medium chain
CC (mu3) and a small chain (sigma3).
CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the
CC cytoplasmic face of coated vesicles located at the Golgi complex
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES LARGE SUBUNITS
CC FAMILY.
CC -!- CAUTION: Sequences in Ref.2 and Ref.3 contain intronic sequences
CC and are incomplete at 5' and 3' ends. Sequences extensively differ
CC from that shown at positions 1-269, 546, and 840-1034.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF002164; AAC14585.1; -  
 DR EMBL; AF044287; AAC01743.1; ALT\_SEQ.  
 DR EMBL; U31351; AAB97618.1; ALT\_SEQ.  
 DR EMBL; AE003493; AAF48307.2; -  
 DR FlyBase; FBgn0001087.5;  
 DR GO; GO:0030123; C:AP-3 adaptor complex; NAS.  
 DR GO; GO:0005783; C:endoplasmic reticulum; IDA.  
 DR GO; GO:0005795; C:Golgi stack; IDA.  
 DR GO; GO:0005798; C:Golgi vesicle; NAS.  
 DR GO; GO:0008057; P:eye pigment granule morphogenesis (sensu Dr. . .; IMP.  
 DR GO; GO:0006985; P:eye pigment granule morphogenesis (sensu Dr. . .; IMP.  
 DR GO; GO:0006727; P:omochromic biosynthesis; IMP.  
 DR InterPro; IPR02553; Adaptin\_N.  
 DR Pfam; PF01602; Adaptin\_N; 1.  
 DR Golgi stack; Protein transport; Transport.  
 DR DOMAIN 767 785 LYS-RICH.  
 FT CONFLICT 395 395 D -> T (IN REF. 2).  
 FT CONFLICT 423 423 L -> V (IN REF. 2).  
 FT CONFLICT 442 445 QILD -> RTTY (IN REF. 2).  
 FT CONFLICT 461 465 MTNLL -> IDQSA (IN REF. 2).  
 FT CONFLICT 694 695 QR -> GO (IN REF. 2).  
 FT CONFLICT 701 701 E -> D (IN REF. 1).  
 FT CONFLICT 869 869 L -> S (IN REF. 4).  
 FT CONFLICT 910 910 P -> S (IN REF. 4).  
 SQ SEQUENCE 1034 AA; 114845 MW; ECE1B7A34DC5FBF1 CRC64;  
 Query Match 2.7%; Score 7; DB 1; Length 1034;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 144 AAVQPV 150  
 Db 906 AAVQPV 912  
 [1]  
 RESULT 33  
 ID 9CA\_BACTO STANDARD; PRT; 1157 AA.  
 AC 045733;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pesticidal crystal protein cry9ca (insecticidal delta-endotoxin  
 DE CryIXC(a)) (crystalline entomocidal protoxin) (130 kDa crystal  
 DE protein).  
 OS CRY9CA OR CRYIXC(A).  
 OC Bacillus thuringiensis (subsp. tolworthi).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 ON NCBI\_TaxID=1442;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=BTS02618A;  
 RX MEDLINE=96141404; PubMed=8572715;  
 RA Lambert B., Buysse L., Decocq C., Jansens S., Pien C., Saey B.,  
 RA Serrinck J., Van Audenheove K., Van Rie J., Van Vleet A., Peteren M.,  
 RT "A Bacillus thuringiensis insecticidal crystal protein with a high  
 RT activity against members of the family Noctuidae.",  
 RL Appl. Environ. Microbiol. 62:80-86(1996).  
 CC -1- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDDIT  
 CC ETHETAL CELLS OF LEPIDOPTERA LARVAE. HAS A FAIRLY BROAD  
 CC SPECTRUM OF ACTIVITY AGAINST MEMBERS OF THE FAMILIES PYRALIDAE,  
 CC PUTEILLIDAE, SPRINGIDAE, AND NOCTUIDAE. IT IS THE FIRST  
 CC INSECTICIDAL CRYSTAL PROTEIN WITH ACTIVITY AGAINST CUTORMS. NO  
 CC ACTIVITY IS OBSERVED AGAINST SOME BEETLES, SUCH AS COLORADO POTATO  
 CC BEETLE.  
 CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING

CC SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART OF  
 CC THE SPORE COAT.  
 CC -1- TERMINAL:US: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE N-  
 CC TERMINUS.  
 CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z37527; CAA85764.1; -  
 DR PIR; A59350; S49247.  
 DR HSRP; P07130; IDLC.  
 DR InterPro; IPR001178; Endotoxin.  
 DR InterPro; IPR005638; endotoxin\_C.  
 DR InterPro; IPR005639; endotoxin\_N.  
 DR Pfam; PF00555; endotoxin\_1.  
 DR Pfam; PF03944; endotoxin\_C; 1.  
 DR Pfam; PF03945; endotoxin\_N; 1.  
 DR Toxin; Sporulation.  
 SQ SEQUENCE 1157 AA; 129775 MW; C364391EF7FDFB8A CRC64;  
 Query Match 2.7%; Score 7; DB 1; Length 1157;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 57 ASAPVSG 63  
 Db 523 ASAPVSG 529  
 [1]  
 RESULT 34  
 ID POLG\_KUNJM STANDARD; PRT; 3433 AA.  
 AC P14335; Q82983;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polypeptide [contains: capsid protein C (core protein); Matrix  
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural  
 DE proteins NS1, NS2, NS4A and NS4B; Protease/helicase (EC 3.4.21.98)  
 DE (NS3); RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)].  
 OS Kunjin virus (strain MRM61C).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 ON NCBI\_TaxID=11078;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86089524; PubMed=282659;  
 RA Cota G., Parker M.D., Speight G., Byrne M.E., Westaway E.G.,  
 RT "Nucleotide and complete amino acid sequences of Kunjin virus:  
 RT definitive gene order and characteristics of the virus-specified  
 RT proteins.",  
 RL J. Gen. Virol. 69:1-21(1988).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; D00246; BAA00176.1; -  
 DR PIR; A28697; GNMVKV.  
 DR HSSP; P14336; 1SVB.  
 DR MEROPS; S07.001; -  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR000336; Flavi\_glycoprote.  
 DR InterPro; IPR001850; Flavi\_helicase.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR001157; Flavi\_NS1.  
 DR InterPro; IPR000752; Flavi\_NS2A.  
 DR InterPro; IPR000487; Flavi\_NS2B.  
 DR InterPro; IPR001528; Flavi\_NS4A.  
 DR InterPro; IPR000208; Flavi\_NS5.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR002877; Ftsu.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_Pstir.  
 DR Pfam; PF01003; Flavi\_capsid\_1.  
 DR Pfam; PF02832; Flavi\_glycop\_C\_1.  
 DR Pfam; PF00869; Flavi\_glycoprot\_1.  
 DR Pfam; PF00949; Flavi\_helicase\_1.  
 DR Pfam; PF01004; Flavi\_M\_1.  
 DR Pfam; PF00948; Flavi\_NS1\_1.  
 DR Pfam; PF01005; Flavi\_NS2A\_1.  
 DR Pfam; PF01002; Flavi\_NS2B\_1.  
 DR Pfam; PF01350; Flavi\_NS4A\_1.  
 DR Pfam; PF00972; Flavi\_NS4B\_1.  
 DR Pfam; PF01570; Flavi\_propep\_1.  
 DR Pfam; PF01728; Ftsu\_1.  
 DR Pfam; PF00271; helicase\_C\_1.  
 DR ProDom; PD001556; Flavi\_glycoprote\_1.  
 DR ProDom; PD001496; Flavi\_NS1\_1.  
 DR SMART; SM00487; DEXDC\_1.  
 DR SMART; SM00490; HELICC\_1.  
 DR PolyProtein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;  
 DR Core protein; Coat protein; Envelope protein; Hydrolyase; Helicase;  
 DR ATP-binding; Transmembrane; Nonstructural protein.  
 DR INIT\_MET 1  
 FT CHAIN 1 123  
 FT PROPEP 124 215  
 FT CHAIN 216 290  
 FT CHAIN 291 791  
 FT CHAIN 792 1143  
 FT CHAIN 1144 1374  
 FT CHAIN 1375 1505  
 FT CHAIN 1506 2124  
 FT CHAIN 2125 2273  
 FT CHAIN 2274 2528  
 FT CHAIN 2529 3433  
 FT DOMAIN 388 401  
 FT NP\_BIND 1699 1706  
 FT SITE 1780 1793  
 FT DISULFID 293 320  
 FT DISULFID 350 406  
 FT DISULFID 364 395  
 FT DISULFID 382 411  
 FT DISULFID 480 578  
 FT DISULFID 595 626  
 FT CAROXYD 138 138  
 FT CAROXYD 921 921  
 FT CAROXYD 966 966  
 FT CAROXYD 998 998

MEMBRANE PROTEIN M.  
 MAJOR ENVELOPE PROTEIN E.  
 NONSTRUCTURAL PROTEIN NS1.  
 NONSTRUCTURAL PROTEIN NS2A.  
 NONSTRUCTURAL PROTEIN NS2B.  
 PROTEASE/HELICASE (NS3).  
 NONSTRUCTURAL PROTEIN NS4A.  
 NONSTRUCTURAL PROTEIN NS4B.  
 RNA-DIRECTED RNA POLYMERASE (NS5).  
 INVOLVED IN FUSION.  
 ATP (POTENTIAL).  
 DEAH BOX.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3433 AA; 381363 MW; EA5B88A7D040B99 CRC64;  
 Query Match 2.7%; Score 7; DB 1; Length 3433;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 PGLRCLN 48  
 DB 1307 PGLRCLN 1313

RESULT 35  
 LAMA\_EMENT  
 ID LAMA\_EMENT STANDARD; PRT; 41 AA.  
 AC P38095;  
 DT 01-OCT-1994 (Rel. 30; Created)  
 DT 01-OCT-1994 (Rel. 30; Last sequence update)  
 DT 01-OCT-1994 (Rel. 30; Last annotation update)  
 DE Lactam utilization protein LAMA (fragment).  
 GN LAMA.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eucotiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=162425;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92107186; PubMed=1729609;  
 RA Richardson I.B., Katz M.E., Hynes M.J.;  
 RT "Molecular characterization of the lam locus and sequences involved  
 RT in regulation by the Ambr protein of Aspergillus nidulans.";  
 RL Mol. Cell. Biol. 12:337-346(1992).  
 CC - FUNCTION: INVOLVED IN THE UTILIZATION OF LACTAMS SUCH AS 2-  
 CC PYROLIDINONE.  
 CC - INDUCTION: BY BETA-ALANINE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M77283; AAA3312.1; ALT\_SEQ.  
 DR PIR; A42064; A42064.  
 DR NON\_TER 41  
 DR SEQUENCE 41 AA; 4458 MW; D9BE04A383A78FF2 CRC64;  
 Query Match 2.3%; Score 6; DB 1; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 TSPARG 43  
 DB 33 TSPARG 38

RESULT 36  
 Y185\_METUA  
 ID Y185\_METUA STANDARD; PRT; 49 AA.  
 AC Q57644;  
 DT 01-NOV-1997 (Rel. 35; Created)  
 DT 01-NOV-1997 (Rel. 35; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE Hypothetical protein M01085.  
 GN M01085.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

```

RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Raulo G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kesteven A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67475; AAB98182.1; -.
DR PIR; B64323; B64323.
DR TIGR; MJ0185; -.
KM Hypothetical protein; Transmembrane, Complete proteome.
FT TRANSMEM 22 42 POTENTIAL.
SQ SEQUENCE 49 AA; 5332 MW; D1594F2AAD47CEFA CRC64;

Query March 2.3%; Score 6; DB 1; Length 49;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 IIAIG 185
Db 30 IIAIG 35

RESULT 37
ID BD02 RAT STANDARD; PRT; 63 AA.
AC O88514;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-defensin 2 precursor (BD-2) (RBD-2).
DE DEFB2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=99386883; PubMed=10456937;
RA Jia H.-P., Mills J.N., Barahmand-Pour F., Nishimura D.,
RA Mallampati R.K., Wang G., Miles K., Tack B.F., Bevins C.L.,
RA McClary P.B., Jr.;
RT "Molecular cloning and characterization of rat genes encoding
RT homologues of human beta-defensins."
RL Infect. Immun. 67:4827-4833(1999).
CC -----
CC -1- FUNCTION: HAS BACTERICIDAL ACTIVITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted (potential).
CC -1- TISSUE SPECIFICITY: Highly expressed in lung.
CC -1- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY. LAP/TAP
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; AF068861; AAC28072.1; -.
DR HSSP; O15263; 1FD3.
DR InterPro; IPR001855; Defensin_beta.
DR InterPro; IPR006080; Defensin_mammal.
DR Pfam; PF00711; Defensin_beta_1.
DR SMART; SM00048; DEFSN; 1.
KW Antibiotic; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 72 POTENTIAL.
FT CHAIN 22 63 BETA-DEFENSIN 2.
FT DISULFID 31 59 BY SIMILARITY.
FT DISULFID 38 52 BY SIMILARITY.
FT DISULFID 42 60 BY SIMILARITY.
SQ SEQUENCE 63 AA; 6946 MW; 826099DE214ACFA CRC64;

Query March 2.3%; Score 6; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 PLSAFT 224
Db 17 PLSAFT 22

RESULT 38
ID CSPE_PSEAE STANDARD; PRT; 69 AA.
AC P95459;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major cold shock protein cspe.
DE CSPE OR PA3266.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales.
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Gollity L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -----
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- INDUCTION: In response to low temperature.
CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U82823; AAB40922.1; -.
DR EMBL; AE004749; AAC06554.1; -.
DR PIR; H83236; H83236.

```

DR HSSP; P32081; 1CSP.  
DR InterPro: IPR002059; Cold\_shock.  
DR Pfam: PF00313; GSD; 1.  
DR PRINTS; PR00050; COLDSHOCK.  
DR ProDom; PD000621; Cold\_shock; 1.  
DR SMART; SM00357; GSD; 1.  
DR PROSITE; PS00352; COLD\_SHOCK; 1.  
KW Transcription regulation; DNA-binding; Activator; Multigene family;  
KM Complete proteome.  
FT DOMAIN 7 66 GSD.  
SQ SEQUENCE 69 AA; 7606 MW; 993225127C41AE43 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 69;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 ADEVQV 130  
Db 63 ADEVQV 68

RESULT 39  
PYS1 SYNCL STANDARD; PRT; 78 AA.  
AC P50035;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Psychobillism 8.9 kDa linker polypeptide, phycocyanin-associated, rod  
DE (Rod capping linker protein).  
GN CPD OR TSR1960.  
OS Synecchococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.  
OX NCBI\_TaxID=32046;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Shimazu T., Soga M., Hirano M., Katoh S.;  
RT "Cloning and sequencing of the phycocyanin operon from the  
RT thermophilic cyanobacterium Synecchococcus elongatus.";  
RT Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=BP-1;  
RX MEDLINE=2225144; PubMed=12240834;  
RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium  
RT Thermosynechococcus elongatus BP-1.";  
RL DNA Res. 9:123-130(2002).  
CC -!- SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D13173; BAA02458.1; -;  
DR EMBL; AF005375; BAC09512.1; -;  
DR InterPro; IPR001685; Cpcd-like.  
DR Pfam; PF01383; Cpcd; 1.  
DR ProDom; PD002828; Cpcd-like\_C; 1.  
KW Psychobillism; Photosynthesis; Complete proteome.  
KW PhycoBillism; Photosynthesis; Complete proteome.  
SQ SEQUENCE 78 AA; 8672 MW; E529978375F30247 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 EMQRT 217  
Db 58 EMQRT 63

RESULT 40  
ATPL MYCLE STANDARD; PRT; 81 AA.  
ID ATPL MYCLE  
AC P45828;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE ATP synthase c chain (EC 3.6.3.14) (lipid-binding protein)  
DE (dicyclohexylcarbodiimide-binding protein).  
GN ATPE OR ML1140.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Smith D.R., Robison K.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Ruter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrett B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RT Nature 409:1007-1011(2001).  
RL Nature 409:1007-1011(2001).  
CC -!- FUNCTION: THIS IS ONE OF THE THREE CHAINS OF THE NONENZYMATIC  
CC COMPONENT (CF(0) SUBUNIT) OF THE ATPASE COMPLEX.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +  
CC H(+) (out).  
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC  
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)  
CC HAS THREE MAIN SUBUNITS: A, B AND C.  
CC -!- MISCELLANEOUS: DICYCLOHEXYLCARBODIIMIDE (DCDI) INHIBITS ATPASE.

CC -!- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U15186; AAA63107.1; -;  
DR EMBL; AB589520; CAC31521.1; -;  
DR PIR; T09979; T09979.  
DR Leproma; ML1140; -;  
DR InterPro; IPR005953; ATP synt C.  
DR InterPro; IPR002379; ATPase Csub.  
DR InterPro; IPR000454; Eub ATPase Csub.  
DR Pfam; PF00137; ATP-synt\_C; 1.  
DR PRINTS; PR00124; ATPaseC.  
DR TIGRfams; TIGR01260; ATP synt C; 1.  
DR PROSITE; PS00605; ATPase\_C; 1.  
KW Hydrogen ion transport; CF(0); Transmembrane; Lipid-binding;  
KW Complete proteome.

FT TRANSMEM 5 25 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT BINDING 61 61 DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).



SQ SEQUENCE 81 AA; 8205 MW; E5456207EE9E0C13 CRC64;  
 Query Match 2.3%; Score 6; DB 1; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 183 AIGAGI 188  
 Db 21 AIGAGI 26  
 RESULT 41  
 ATPL\_MYCTU STANDARD; PRT; 81 AA.  
 ID Q10598;  
 AC Q10598;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ATP synthase C chain (EC 3.6.3.14) (Lipid-binding protein)  
 DE (Dicyclohexylcarbodiimide-binding protein).  
 GN ATP OR RV1305 OR MT1345 OR MTCY373.25.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxId=1773;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Bignlmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,  
 RT "Deciphering the Biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Petersmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gilm M.L., Hatt D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS IS ONE OF THE THREE CHAINS OF THE NONENZYMATIC  
 CC COMPONENT (CF(0) SUBUNIT) OF THE ATPASE COMPLEX.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +  
 CC H(+) (out).  
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC  
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), Epsilon(1). CF(0)  
 CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: DICYLOHEXYLCARBODIIMIDE (DCCD) INHIBITS ATPASE.  
 CC -1- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z73419; CAA97738.1; -  
 CC EMBL; AE007008; AAK45607.1; -  
 CC PIR; E70774; E70774.

DR TIGR; MT1345; -  
 DR TubercuList; RV1305; -  
 DR InterPro; IPR005953; ATP\_synt\_C.  
 DR InterPro; IPR002379; ATPase\_Csub.  
 DR InterPro; IPR000454; Eub ATPase\_Csub.  
 DR Pfam; PF00137; ATP-synt\_C; 1.  
 DR PRINTS; PR00124; ATPASEC.  
 DR TIGRfam; TIGR01260; ATP\_synt\_c; 1.  
 DR PROSITE; PS00605; ATPASE\_C; 1.  
 KM Hydrogen ion transport; CF(0); Transmembrane; Lipid-binding;  
 KM Complete proteome.  
 FT TRANSMEM 5 25  
 FT TRANSHEM 57 77  
 FT BINDING 61 61  
 FT DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).  
 SQ SEQUENCE 81 AA; 8055 MW; 36303496CDDF52CC CRC64;  
 Query Match 2.3%; Score 6; DB 1; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 183 AIGAGI 188  
 Db 21 AIGAGI 26  
 RESULT 42  
 ID RS17\_CHLMU STANDARD; PRT; 83 AA.  
 AC Q9PUM3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30S ribosomal protein S17.  
 GN RPSO OR TC0806.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxId=93560;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey B.K., Peterson J., Uitterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gilm M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR35.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 CC -1- FUNCTION: PROTEIN S17 BINDS SPECIFICALLY TO THE 5' END OF 16S  
 CC RIBOSOMAL RNA (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE S17P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE002347; AAF39609.1; -  
 CC PIR; F81664; F81664.  
 CC HSSP; P23828; IRIp.  
 CC TIGR; TC0806; -  
 DR InterPro; IPR00266; Ribosomal\_S17.  
 DR Pfam; PF00366; Ribosomal\_S17; 1.  
 DR PRINTS; PR00973; RIBOSOMALS17.  
 DR ProDom; PD001295; RIBOSOMAL\_S17; 1.  
 DR PROSITE; PS00056; RIBOSOMAL\_S17; 1.  
 KM Ribosomal protein; rRNA-binding; Complete proteome.  
 SQ SEQUENCE 83 AA; 9703 MW; 446C85F64418FD2 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 EKTUVV 238  
 DB 22 EKTUVV 27

RESULT 43

RS17\_CHLTP  
 ID RS17\_CHLTP STANDARD; PRT; 83 AA.  
 AC P28545;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30S ribosomal protein S17.  
 GN RPSQ OR RS17 OR CTS19.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=L2/434/Bu;  
 RX MEDLINE=92138612; PubMed=1735714;  
 RA Kaul R., Gray G.J., Koehnke N.R., Gu L.J.,  
 RT "Cloning and sequence analysis of the Chlamydia trachomatis spc  
 RT ribosomal protein gene cluster";  
 RL J. Bacteriol. 174:1205-1212(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=D/UV-3/Cx;  
 RX MEDLINE=9900809; PubMed=9784136;  
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.,  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis";  
 RL Science 282:754-759(1998).  
 CC -1- FUNCTION: PROTEIN S17 BINDS SPECIFICALLY TO THE 5' END OF 16S  
 CC RIBOSOMAL RNA (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE S17P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M80325; AAA23171.1; -  
 CC EMBL; AE001323; AAC68120.1; -  
 CC PIR; C42645; C42645.  
 CC HSSP; P23828; 1RIP.  
 CC InterPro; IPR000266; Ribosomal\_S17.  
 CC Pfam; PF00366; Ribosomal\_S17.1.  
 CC PRINTS; PR00973; RIBOSOMALS17.  
 CC ProDom; PD001295; RIBOSOMAL\_S17.1.  
 CC PROSITE; PS00056; RIBOSOMAL\_S17.1.  
 CC Riboosomal protein; rRNA-binding; Complete proteome.  
 CC SEQUENCE 83 AA; 9645 MW; 446CB5FDB418FD2 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 EKTUVV 238  
 DB 22 EKTUVV 27

RESULT 44

RS17\_CHLTP  
 ID RS17\_CHLTP STANDARD; PRT; 86 AA.  
 AC Q92786; Q9J0G6;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30S ribosomal protein S17.  
 GN RPSQ OR RS17 OR CPN0638 OR CP0109.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Knout H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.,  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.,  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CML029 from USA";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- FUNCTION: PROTEIN S17 BINDS SPECIFICALLY TO THE 5' END OF 16S  
 CC RIBOSOMAL RNA (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE S17P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE01647; AAD18777.1; -  
 CC EMBL; AE002173; AAF37992.1; -  
 CC EMBL; AP002547; BAA9845.1; -  
 CC PIR; C86570; C86570.  
 CC PIR; H72054; H72054.  
 CC HSSP; F23628; 1RIP.  
 CC TIGR; CP0109; -  
 CC InterPro; IPR000266; Ribosomal\_S17.  
 CC Pfam; PF00366; Ribosomal\_S17.1.  
 CC PRINTS; PR00973; RIBOSOMALS17.  
 CC ProDom; PD001295; RIBOSOMAL\_S17.1.  
 CC PROSITE; PS00056; RIBOSOMAL\_S17.1.  
 CC Riboosomal protein; rRNA-binding; Complete proteome.  
 CC SEQUENCE 86 AA; 9889 MW; 7EA2542922FC14 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 EKTUVV 238  
 DB 22 EKTUVV 27

```

Db          22 EKTVVV 27

RESULT 45
ID IFHB_PASHA STANDARD; PRT; 93 AA.
AC P95519;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integration host factor beta-subunit (IHf-beta).
GN IHFB OR HIMD.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Mannheimia.
OX NCBI_TaxID=75985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype A1 / PH1101;
RX MEDLINE=97164347; PubMed=9011038;
RA Highlander S.K., Garza O., Brown B.J., Koby S., Oppenheim A.B.;
RT "Isolation and characterization of the integration host factor genes
of Pasteurella haemolytica."
RL FEMS Microbiol. Lett. 146:181-188 (1997).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE TWO SUBUNITS OF INTEGRATION
CC HOST FACTOR. A SPECIFIC DNA-BINDING PROTEIN THAT FUNCTIONS IN
CC GENETIC RECOMBINATION AS WELL AS IN TRANSCRIPTIONAL AND
CC TRANSLATIONAL CONTROL (BY SIMILARITY).
CC -1- SUBUNIT: Heterodimer of an alpha and a beta chain.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U56139; AAC44846.1; -.
DR HSSP; P36206; 1B82.
DR HAMAP; MF_00381; -. 1.
DR InterPro; IPR000119; Bac_DNAbind.
DR InterPro; IPR005685; IHF_beta.
DR Pfam; PF00216; Bac_DNA_binding; 1.
DR ProDom; PD000945; Bac_DNAbind; 1.
DR SMART; SM00411; BHL; 1.
DR TIGRFSMS; TIGR00988; hfp; 1.
DR PROSITE; PS00045; HISTONE_LIKE; 1.
KW DNA-binding; Transcription regulation; DNA recombination;
KW Translation regulation.
SQ SEQUENCE 93 AA; 10490 MW; 7E29C35CE4D4A6C CRC64;

Query Match 2.3%; Score 6; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 GKDLKE 202
Db 83 GKDLKE 88

RESULT 46
ID Y492_METUA STANDARD; PRT; 95 AA.
AC O57915;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0492.
GN MJ0492.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

Db          22 EKTVVV 27

OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96377999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.T.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uitterback L.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurest M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL Science 273:1058-1073 (1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U67499; AAB98488.1; -.
DR PIR; D64361; D64361.
DR TIGR; MJ0492; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
PT TRANSMEM 3 23 POTENTIAL.
SQ SEQUENCE 95 AA; 10694 MW; 7A49FD4F888A365C CRC64;

Query Match 2.3%; Score 6; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 VIIIAI 184
Db 4 VIIIAI 9

RESULT 47
ID THIO_ALIAC STANDARD; PRT; 105 AA.
AC P80579;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thioredoxin (TRX).
GN TRXA.
OS Alicyclobacillus acidocaldarius (Bacillus acidocaldarius).
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae;
OC Alicyclobacillus.
OX NCBI_TaxID=1388;
RN [1]
RP SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=98028685; PubMed=9359865;
RA Barolucci S., Guagliardi A., Pedone E., de Pascale D., Cammio R.,
RA Camardella L., Rossi M., Nicastro G., de Chiara C., Facci P.,
RA Mascetti G., Nicolini C.;
RT "Thioredoxin from Bacillus acidocaldarius: characterization,
RT high-level expression in Escherichia coli and molecular modelling."
RL Biochem. J. 328:277-285 (1997).
RN [2]
RP STRUCTURE BY NMR.
RX MEDLINE=20098516; PubMed=10632710;
RA Nicastro G., De Chiara C., Pedone E., Taro M., Rossi M.,
RA Barolucci S.;
RT "NMR solution structure of a novel thioredoxin from bacillus
RT acidocaldarius: possible determinants of protein stability."
RL Eur. J. Biochem. 267:403-413 (2000).
CC -1- FUNCTION: Participates in various redox reactions through the

```

```
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions.
CC -1- MASS SPECTROMETRY: MW=1157; METHOD=Electrospray.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
DR PDB: 1QW; 26-JAN-00.
DR InterPro: IPR006662; Thiorod.
DR InterPro: IPR006663; Thiorodex dom2.
DR InterPro: IPR005746; Thiorodexin.
DR Pfam: PF00085; Thiorod. 1.
DR PRINTS: PR00421; THIOREDOXIN.
DR TIGRFS: TIGR01068; thiorodexin; 1.
DR PROSITE: PS00194; THIOREDOXIN; 1.
DR Redox-active center; Electron transport; 3D-structure.
FT DISULFID 29 32 REDOX-ACTIVE.
FT HELIX 7 14
FT TURN 15 17
FT STRAND 20 25
FT TURN 27 28
FT TURN 31 31
FT HELIX 32 45
FT TURN 46 46
FT STRAND 52 56
FT TURN 57 59
FT HELIX 62 67
FT TURN 68 68
FT STRAND 74 78
FT STRAND 83 88
FT HELIX 93 105
SQ SEQUENCE 105 AA; 11576 MW; E03F636DFB3C3745 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 PETTSQ 107
DB 61 PETTSQ 66

RESULT 48
RLA3 YEAST STANDARD; PRT; 106 AA.
AC P10622;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60S acidic ribosomal protein P1-beta (L44') (L12E11B).
GN RPL18B OR RPLA3 OR L12E11B OR RPL44P OR YDL130W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88243786; PubMed=2837476;
RA Remacha M., Saenz-Robles M.T., Vilella M.D., Ballesta J.P.G.;
RT "Independent genes coding for three acidic proteins of the large
RT ribosomal subunit from Saccharomyces cerevisiae.";
RL J. Biol. Chem. 263:9094-9101(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=SR26-12C; PubMed=2404943;
RA Newton C.H., Shimmin L.C., Yee J., Dennis P.P.;
RT "A family of genes encode the multiple forms of the Saccharomyces
RT cerevisiae ribosomal proteins equivalent to the Escherichia coli L12
RT protein and a single form of the L10-equivalent ribosomal protein.";
RL J. Bacteriol. 172:579-588(1990).
RN [3]
RP SEQUENCE FROM N.A.
RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M., Wagner G.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
```

```
CC PROTEIN SYNTHESIS.
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -1- MISCELLANEOUS: YEASTS CONTAIN 4 INDIVIDUAL SMALL RIBOSOMAL A
CC PROTEINS (RPA) WHICH CAN BE CLASSIFIED INTO TWO COUPLES OF SIMILAR
CC BUT NOT IDENTICAL SEQUENCES. EACH COUPLE IS DISTINCTLY RELATED TO
CC ONE OF THE TWO A PROTEINS PRESENT IN MULTICELLULAR ORGANISMS.
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19238; AAA34973.1; -
DR EMBL: M26507; AAA34734.1; -
DR EMBL: Z74178; CAA98698.1; -
DR PIR: C28104; R8BY2B.
DR SGD: S0002288; RPL18B.
DR InterPro: IPR001813; 60S_ribosomal.
DR Pfam: PF00428; 60S_ribosomal; 1.
DR Ribosomal protein; Phosphorylation, Multigene family.
KM RIBOSOMAL PROTEIN; 10667 MW; EAED4F748653E0DC CRC64;
SQ SEQUENCE 106 AA; 10667 MW; EAED4F748653E0DC CRC64;

Query Match 2.3%; Score 6; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 GKDIKE 202
DB 50 GKDIKE 55

RESULT 49
THCC_RHOER STANDARD; PRT; 106 AA.
AC P43493;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rhodococcin.
GN Rhodococcin.
OS Rhodococcus erythropolis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-7.
RC STRAIN=N186/21;
RX MEDLINE=95138028; PubMed=7836301;
RA Nagy I., Schoofs G., Compenolle F., Proost P., Vanderleyden J.,
RA de Mot R.;
RT "Degradation of the thiocarbamate herbicide EPTC (S-ethyl
RT diisopropylcarbamothioate) and biosurfing by Rhodococcus sp. strain
RT N186/21 involve an inducible cytochrome P-450 system and aldehyde
RT dehydrogenase.";
RL J. Bacteriol. 177:676-687(1995).
RN [2]
RP FUNCTION: FERREDOXIN-TYPE PROTEIN WHICH TRANSFERS ELECTRONS FROM
RP RHODOCOCCIN REDUCTASE TO CYTOCHROME CYP116 (THCB), WHICH IS
RP INVOLVED IN THE DEGRADATION OF THIOCARBAMATE HERBICIDES.
CC -1- CORRELATOR: BINDS 1 2PE-2S CLUSTER (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ADRENODOXIN / PUTIDAREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC -----
DR EMBL; U17130; AAC45751.1; -.
DR HSSP; P00259; 1GPX.
DR InterPro; IPR001055; Adrenodoxin.
DR InterPro; IPR001041; Ferredoxin.
DR Pfam; PF00111; fer2; 1.
DR PRINTS; PR00355; ADRENODOXIN.
DR PROSITE; PS00814; ADX; 1.
KM Electron transport; Metal-binding; Iron-sulfur; Iron; 2Fe-2S.
FT INIT MET 0
FT METAL 39 39 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 45 45 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 48 48 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 86 86 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 106 AA; 11415 MW; A65AAB91CD5710B CRC64;

Query Match 2.3%; Score 6; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 AIGANGI 188
Db 27 AIGANGI 32

RESULT 50
DBH BORAF STANDARD; PRT; 108 AA.
AC Q44625;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-binding protein HBBU.
GN HBB.
OS Borrelia andersonii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=42109;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=19952;
RA Valsangiacomo C., Balmelli T., Piffaretti J.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48686; AAC73108.1; -.
DR HSSP; P02346; 1HUH.
DR InterPro; IPR000119; Bac DNABind.
DR Pfam; PF00216; Bac_DNA_binding; 1.
DR ProDom; PD000945; Bac_DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE_LIKE; FALSE_NEG.
KM DNA-binding; DNA condensation.
SQ SEQUENCE 108 AA; 12742 MW; E25A6F802B361B88 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 GKDLKE 202
Db 96 GKDLKE 101
```

```
RESULT 51
DBH BORAF STANDARD; PRT; 108 AA.
ID DBH BORAF
AC Q57220; Q44624;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein HBBU.
GN HBB.
OS Borrelia afreilli.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=29518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A26S, B023, DK8, ECML, UO1, and V6461;
RA Valsangiacomo C., Balmelli T., Piffaretti J.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48671; AAC73093.1; -.
DR EMBL; U48672; AAC73094.1; -.
DR EMBL; U48673; AAC73095.1; -.
DR EMBL; U48674; AAC73096.1; -.
DR EMBL; U48675; AAC73097.1; -.
DR EMBL; U48676; AAC73098.1; -.
DR HSSP; P02346; 1HUH.
DR InterPro; IPR000119; Bac DNABind.
DR Pfam; PF00216; Bac_DNA_binding; 1.
DR ProDom; PD000945; Bac_DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE_LIKE; FALSE_NEG.
KM DNA-binding; DNA condensation.
FT VARIANT 19 19 S -> A (IN STRAIN UO1).
FT VARIANT 23 23 R -> K (IN STRAIN UO1).
FT VARIANT 89 89 H -> N (IN STRAIN UO1).
SQ SEQUENCE 108 AA; 12724 MW; 8E96BE73E40AEB41 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 GKDLKE 202
Db 96 GKDLKE 101

RESULT 52
DBH BORBU
ID DBH BORBU STANDARD; PRT; 108 AA.
AC Q57267; Q44834; Q44835; Q44836; Q44837; Q44838; Q44839; Q57030;
AC Q57056; Q57402;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein HU (HBBU).
GN HUP OR HBB OR BB0232.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
```

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9319731; Pubmed=8451174;  
 RA Tilly K., Campbell J.  
 RT "A *Borrelia burgdorferi* homolog of the *Escherichia coli* rho gene."  
 RN Nucleic Acids Res. 21:1040-1040(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Various strains;  
 RA Valsangiaco C., Balmelli T., Piffaretti J.C.;  
 RL Submitted (Feb-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RX MEDLINE=98065943; Pubmed=9403685;  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
 RA Lathigra R., Tomb J.F., Fleischmann R.D., Richardson D.,  
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
 RA Uterback T., Watthey L., McDonald L., Arlatch P., Bowman C.,  
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
 RA Smith H.O., Venter J.C.;  
 RT "Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*."  
 RL Nature 390:580-586(1997).  
 CC -1- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF  
 CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING  
 CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME  
 CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U35673; AAB41461.1; -  
 DR EMBL; U35673; AAC73070.1; -  
 DR EMBL; U48649; AAC73071.1; -  
 DR EMBL; U48650; AAC73072.1; -  
 DR EMBL; U48651; AAC73073.1; -  
 DR EMBL; U48652; AAC73074.1; -  
 DR EMBL; U48653; AAC73075.1; -  
 DR EMBL; U48654; AAC73076.1; -  
 DR EMBL; U48653; AAC73085.1; -  
 DR EMBL; U48654; AAC73087.1; -  
 DR EMBL; U48655; AAC73088.1; -  
 DR EMBL; U48656; AAC73089.1; -  
 DR EMBL; U48667; AAC73090.1; -  
 DR EMBL; U48668; AAC73091.1; -  
 DR EMBL; U48669; AAC73092.1; -  
 DR EMBL; U48670; AAC73092.1; -  
 DR EMBL; U48683; AAC73105.1; -  
 DR EMBL; U48684; AAC73106.1; -  
 DR EMBL; U48685; AAC73107.1; -  
 DR EMBL; A8001133; AAC66617.1; -  
 DR TIGR; BB0232; -  
 DR InterPro; IPR000119; Bac DNAbind.  
 DR Pfam; PF00216; Bac\_DNA\_binding; 1.  
 DR ProDom; PD000945; Bac\_DNAbind; 1.  
 DR PROSITE; PS00045; HISTONE\_LIKE; FALSE NEG.  
 KW DNA-binding; DNA condensation; Complete proteome.  
 FT VARIANT 4 4 S -> P (IN STRAINS UK AND VS116).  
 FT VARIANT 9 9 V -> I (IN STRAINS UK AND VS116).  
 FT VARIANT 14 16 IVD -> VVN (IN STRAINS POT1B1, POT1B2 AND  
 FT POT1B3).  
 FT VARIANT 19 19 S -> A (IN STRAINS A44S AND NT13-87).  
 FT VARIANT 23 23 K -> R (IN STRAINS 25015, CA2, CA55,  
 FT CA128, DN127, POT1B1, POT1B2, POT1B3, UK  
 FT AND VS116).

FT VARIANT 33 33 Y -> C (IN STRAINS CA55, CA128 AND  
 FT DN127).  
 FT VARIANT 65 65 V -> L (IN STRAIN 25015).  
 FT VARIANT 69 69 K -> R (IN STRAIN DN127).  
 FT VARIANT 81 81 E -> V (IN STRAIN NT13-87).  
 SQ SEQUENCE 108 AA; 12686 MW; 0CEP9A092B24BED9 CRC64;  
 Query Match 2.3%; Score 6; DB 1; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 197 GKD1KE 202  
 Db 96 GKD1KE 101  
 RESULT 53  
 ID DBI\_BORCA STANDARD; PRT; 108 AA.  
 AC Q57153; Q45164; Q57235;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-binding protein HBBU.  
 GN HBB.  
 OS *Borrelia garinii*.  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; *Borrelia*.  
 OX NCBI\_TaxId=29519;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A19S, IP99, NT29, PBI, SIKAI, SIKAZ, VS102, and VSBP;  
 RA Valsangiaco C., Balmelli T., Piffaretti J.C.;  
 RL Submitted (Feb-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF  
 CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING  
 CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME  
 CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U48656; AAC73078.1; -  
 DR EMBL; U48657; AAC73079.1; -  
 DR EMBL; U48664; AAC73086.1; -  
 DR EMBL; U48658; AAC73080.1; -  
 DR EMBL; U48659; AAC73081.1; -  
 DR EMBL; U48660; AAC73082.1; -  
 DR EMBL; U48661; AAC73083.1; -  
 DR EMBL; U48662; AAC73084.1; -  
 DR HSSP; P02346; 1HUE.  
 DR InterPro; IPR000119; Bac DNAbind.  
 DR Pfam; PF00216; Bac\_DNA\_binding; 1.  
 DR ProDom; PD000945; Bac\_DNAbind; 1.  
 DR SMART; SM00411; BHL; 1.  
 DR PROSITE; PS00045; HISTONE\_LIKE; FALSE NEG.  
 KW DNA-binding; DNA condensation.  
 FT VARIANT 19 20 SL -> FF (IN STRAINS SIKAI AND SIKAZ2).  
 FT VARIANT 90 90 V -> I (IN STRAIN VS102).  
 SQ SEQUENCE 108 AA; 12724 MW; 4A95CE939FBB57A5 CRC64;  
 QY 197 GKD1KE 202  
 Db 96 GKD1KE 101  
 Query Match 2.3%; Score 6; DB 1; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 54
DBH_BORJA STANDARD; PRT: 108 AA.
ID DBH_BORJA Q45227; Q45228; Q45229; Q45230;
AC Q45231; Q45227; Q45228; Q45229; Q45230;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein HBBU.
GN HBB.
OS Borrelia japonica.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=34095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0612, H014, F63B, COM611A, and COM611C;
RA Valsangiacomo C., Balmelli T., Piffaretti J.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48677; AAC73099.1; -
DR EMBL; U48678; AAC73100.1; -
DR EMBL; U48679; AAC73101.1; -
DR EMBL; U48680; AAC73102.1; -
DR EMBL; U48681; AAC73103.1; -
DR HSSP; P02346; 1HUU.
DR InterPro; IPR000119; Bac_DNABind.
DR Pfam; PF00216; Bac_DNA_Binding; 1.
DR ProDom; PD000945; Bac_DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE_LIKE; FALSE_NEG.
KW DNA-binding; DNA condensation.
SQ SEQUENCE 108 AA; 12684 MW; 90BB42802B220F9E CRC64;

Query Match 2.3%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1,le+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 GKDLKE 202
Db 96 GKDLKE 101

RESULT 55
DBH_BORJU STANDARD; PRT: 108 AA.
ID DBH_BORJU Q45722;
AC Q45722;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-binding protein HBBU.
GN HBB.
OS Borrelia turicatae.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=142;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0M2007;
RA Valsangiacomo C., Balmelli T., Piffaretti J.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48682; AAC73104.1; -
DR HSSP; P02346; 1HUU.
DR InterPro; IPR000119; Bac_DNABind.
DR Pfam; PF00216; Bac_DNA_Binding; 1.
DR ProDom; PD000945; Bac_DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE_LIKE; FALSE_NEG.
KW DNA-binding; DNA condensation.
SQ SEQUENCE 108 AA; 12676 MW; 6B91C5E37A72BD59 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1,le+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 GKDLKE 202
Db 96 GKDLKE 101

RESULT 56
RLAI_SCHPO STANDARD; PRT: 109 AA.
ID RLAI_SCHPO AC P17476;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein P1-alpha 1 (A1).
GN RPA1 OR SPAC644.15.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90220620; PubMed=2325655;
RA Beltrame M., Bianchi M.E.;
RT "A gene family for acidic ribosomal proteins in Schizosaccharomyces
RT pombe: two essential and two nonessential genes.";
RL Mol. Cell. Biol. 10:2341-2348(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94252568; PubMed=8194753;
RA Jang Y.K., Jin Y.H., Kim E.M., Hong S.H., Fabre F., Park S.D.;
RT "Cloning and sequence analysis of rps1+, a Schizosaccharomyces pombe
RT homolog of the Saccharomyces cerevisiae Rps1 gene.";
RL Gene 142:207-211(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=2184401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham S., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsen K.,
RA James K., Jones L., Jones W., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,

```





Query Match 2.3%; Score 6; DB 1; Length 109;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 GKDLKE 202  
Db 52 GKDLKE 57

RESULT 58  
RS17\_HALN1 STANDARD; PRT; 109 AA.  
AC 024786; G9HPC5;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S17 (HHAS17).  
GN RPS17P OR VNG1700G.  
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081), and  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium halobium.  
OX NCBI\_TaxID=64091, 2242;  
[1]  
[2]  
SEQUENCE FROM N.A.  
RA MEDLINE=20504483; PubMed=11016950;  
RA NG W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
RA Shukla H.D., Laskey S.R., Balliga N.S., Thorson V., Sproga J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir A.,  
RA Lettner E., Keller K., Cruz R., Danson M.J., Hough D.M., Dale H.,  
RA Maddocks D.G., Jablonski P.E., Krebs W.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlischroder M., Spudich J.L., Jung K.-H.,  
RA Alam M., Freltas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
RT "Genome sequence of Halobacterium species NRC-1";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
[2]  
SEQUENCE FROM N.A.  
RA SPECIES=H.halobium;  
RC MEDLINE=97031049; PubMed=8876975;  
RA Miyokawa T., Urayama T., Shimooka K., Itoh T.;  
RT "Organization and nucleotide sequences of ten ribosomal protein genes  
from the region equivalent to the S10 operon in the archaeobacterium,  
RT Halobacterium halobium";  
RL Biochem. Mol. Biol. Int. 39:1209-1220(1996).  
CC -1- SIMILARITY: BELONGS TO THE S17P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; AB005077; AAG19945.1; -;  
DR EMBL; AB006961; BAA22279.1; -;  
DR PIR; E84322; E84322.  
DR PIR; T43825; T43825.  
DR HSSP; P23828; TRIP.  
DR InterPro; IPR000266; Ribosomal\_S17.  
DR Pfam; PF00366; Ribosomal\_S17; 1.  
DR PRINTS; PR00973; RIBOSOMAL\_S17.  
DR ProDom; PD001295; Ribosomal\_S17; 1.  
DR PROSITE; PS00056; RIBOSOMAL\_S17; 1.  
KW Ribosomal protein; rRNA-binding; Complete proteome.  
SQ SEQUENCE 109 AA; 11973 MW; E8B9B80365610A3C CRC64;

Query Match 2.3%; Score 6; DB 1; Length 109;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 EKTIVV 238  
Db 42 EKTIVV 47

RESULT 59  
RLA3\_SCHPO STANDARD; PRT; 110 AA.  
AC P17477;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 60S acidic ribosomal protein P1-alpha 3 (A3).  
GN RPA3 OR SPB3B9.13C.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
[1]  
[2]  
SEQUENCE FROM N.A.  
RA MEDLINE=90220620; PubMed=2325655;  
RA Beltrame M., Bianchi M.E.;  
RT "A gene family for acidic ribosomal proteins in Schizosaccharomycetes  
RT pombe: two essential and two nonessential genes.";  
RL Mol. Cell. Biol. 10:2341-2348(1990).  
[2]  
SEQUENCE FROM N.A.  
RA STRAIN=972;  
RC  
RX MEDLINE=21844401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Voicikert G., Aert R., Robben J., Gymnopoulos B.,  
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
RA Egger P., Zimmermann W., Medler H., Mambutt R., Punnett B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Domiguez L., Revuelta J.B., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Porashkin J.,  
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomycetes pombe";  
RL Nature 415:871-880(2002).  
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF  
CC PROTEIN SYNTHESIS.  
CC  
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL  
CC SUBUNIT.  
CC  
CC -1- MISCELLANEOUS: YEASTS CONTAIN 4 INDIVIDUAL SMALL RIBOSOMAL A  
CC PROTEINS (RPA) WHICH CAN BE CLASSIFIED INTO TWO COUPLES OF SIMILAR  
CC BUT NOT IDENTICAL SEQUENCES. EACH COUPLE IS DISTINCTLY RELATED TO  
CC ONE OF THE TWO A PROTEINS PRESENT IN MULTICELLULAR ORGANISMS.  
CC  
CC -1- MISCELLANEOUS: RPA3 AND RPA4 ARE ESSENTIAL FOR CELL SURVIVAL,  
CC WHEREAS RPA1 AND RPA2 ARE NOT.  
CC  
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M3139; AAA3536.1; -  
 DR EMBL; AL022070; CA117793.1; -  
 DR PIR; C34715; R6BYP3.  
 DR GenDB Spombe; SPBC3B9.13G; -  
 DR InterPro; IPR001813; 60s\_ribosomal.  
 DR Pfam; PF00428; 60s\_ribosomal; 1.  
 DR Ribosomal protein; Phosphorylation; Multigene family.  
 KW SEQUENCE 110 AA; 11171 MW; 1734AC9779F5A891 CRC64;  
 SQ

Query Match  
 Best Local Similarity 2.3%; Score 6; DB 1; Length 110;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 GKDLKE 202  
 DB 52 GKDLKE 57

RESULT 60  
 RS17 HALMA STANDARD; PRT; 111 AA.  
 ID RS17 HALMA  
 AC P12741;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 30S ribosomal protein S17 (HmsA17) (Hs14).  
 GN RS17P.  
 OS Halococcus marismortui (Halobacterium marismortui).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halocaula.  
 OX NCBI\_TaxID=2238;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90336772; PubMed=2143141;  
 RA Ande E.;  
 RT "Nucleotide sequence of four genes encoding ribosomal proteins from  
 RT the 'S10 and spectinomycin' operon equivalent region in the  
 RT archaebacterium Halobacterium marismortui.";  
 RL FEMS Lett. 267:193-198(1990).  
 RN [2]  
 RP SEQUENCE.  
 RX MEDLINE=87308217; PubMed=3305503;  
 RA Kimura J., Kimura M.;  
 RT "The primary structures of ribosomal proteins S14 and S16 from the  
 RT archaebacterium Halobacterium marismortui. Comparison with  
 RT eubacterial and eukaryotic ribosomal proteins.";  
 RL J. Biol. Chem. 262:12150-12157(1987).  
 CC -1- SIMILARITY: BELONGS TO THE S17P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; X55311; CA139017.1; -  
 DR PIR; S10733; R3HS17.  
 DR HSSP; P23828; IRIP.  
 DR InterPro; IPR000266; Ribosomal\_S17.  
 DR Pfam; PF00366; Ribosomal\_S17; 1.  
 DR PRINTS; PR00973; RIBOSOMAL\_S17.  
 DR PRODOM; PD001295; Ribosomal\_S17; 1.  
 DR PROSITE; PS00056; RIBOSOMAL\_S17; 1.  
 KW Ribosomal protein; rRNA-binding.  
 FT INIT MET 0  
 FT CONFLICT 88 88 C -> S (IN REF. 2).  
 FT CONFLICT 92 92 S -> P (IN REF. 2).  
 CC

FT CONFLICT 109 110 MISSING (IN REF. 2).  
 SQ SEQUENCE 111 AA; 12141 MW; C22DD05891C383A1 CRC64;  
 SQ

Query Match  
 Best Local Similarity 2.3%; Score 6; DB 1; Length 111;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 EKTUVV 238  
 DB 41 EKTUVV 46

RESULT 61  
 IAA4\_SORBI STANDARD; PRT; 118 AA.  
 ID IAA4\_SORBI  
 AC P81367;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Alpha-amylase inhibitor 4 (SI alpha-4).  
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACCAD clade; Panicoideae; Andropogonae; Sorghum.  
 OX NCBI\_TaxID=4558;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=cv. French red; TISSUE=Seed;  
 RX MEDLINE=93149996; PubMed=1492093;  
 RA Bloch C. Jr., Richardson M.;  
 RT "The amino acid sequences of two 13-kDa alpha-amylase inhibitors from  
 RT the seeds of Sorghum bicolor (L.) Moench.";  
 RT Protein Seq. Data Anal. 5:27-30(1992).  
 CC -1- FUNCTION: ALPHA-AMYLASE INHIBITOR.  
 CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPsin/ALPHA-AMYLASE INHIBITOR  
 CC FAMILY.  
 CC -----  
 DR HSSP; P01086; 1BEA.  
 DR InterPro; IPR003612; AAI.  
 DR InterPro; IPR006106; Amylase\_inhib.  
 DR InterPro; IPR006105; Try/amyL\_inhib.  
 DR Pfam; PF00234; tryp\_alpha\_ami1; 1.  
 DR PRINTS; PR00808; AMYLASEINHTR.  
 DR SMART; SM00499; AAI; 1.  
 DR PROSITE; PS00426; CEREAL\_TRYP\_AMYL\_INH; 1.  
 KW Alpha-amylase inhibitor.  
 FT DISULFID 7 60 BY SIMILARITY.  
 FT DISULFID 21 49 BY SIMILARITY.  
 FT DISULFID 30 82 BY SIMILARITY.  
 FT DISULFID 50 101 BY SIMILARITY.  
 SQ SEQUENCE 118 AA; 12499 MW; B957110591806BF CRC64;  
 SQ

Query Match  
 Best Local Similarity 2.3%; Score 6; DB 1; Length 118;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 REMQRI 216  
 DB 84 REMQRI 89

RESULT 62  
 YHAH\_ECOLI STANDARD; PRT; 121 AA.  
 ID YHAH\_ECOLI  
 AC P42621;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yhaH.  
 DR YHAH OR B3103 OR Z4457 OR ECS3985.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.

```

OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RA MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Weich R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RA MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Ikeda T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG, TO E. COLI YHAI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18997; AAA57907.1; ALT_FRAME.
DR EMBL; AE000392; AAC76138.1; -.
DR EMBL; AE005540; AAG58236.1; -.
DR EMBL; AP002564; BAB37408.1; -.
DR PIR; A91127; A91127.
DR PIR; D65099; D65099.
DR PIR; H85971; H85971.
DR ECGene; EG12747; YhaH.
KM Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
SQ SEQUENCE 121 AA; 14281 MW; 6D1D3283E025061C CRC64;

Query Match 2.3%; Score 6; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPL7.
OS Liberibacter asiaticus (Liberibacter asiaticum).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Candidatus Liberibacter.
OX NCBI_TaxID=34021;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9153297; PubMed=7763375;
RA Vilhechanoux S., Garnier M., Laigret F., Renaudin J., Bove J.M.;
RT "The genome of the non-cultured, bacterial-like organism associated
RT with citrus greening disease contains the nusG-rplKvL-rpoC gene
RT cluster and the gene for a bacteriophage type DNA polymerase.",
RL Curr. Microbiol. 26:161-166 (1993).
CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC ACCURATE TRANSLATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M94319; AAA2108.1; -.
DR HSSP; P03392; ICTF.
DR HAMAP; MF_00368; -.
DR InterPro; IPR000206; Ribosomal L12.
DR Pfam; PF00542; Ribosomal L12; I.
DR ProDom; PD001326; Ribosomal L12; 1.
DR TrRfam; TRFR0085; L12; 1.
KM Ribosomal protein.
SQ SEQUENCE 122 AA; 12955 MW; FC30A0AAC8D2926E CRC64;

Query Match 2.3%; Score 6; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 57 ASAPVS 62
ID [1]
AC 34 ASAPVS 39
DB 34 ASAPVS 39
ID [1]
ID NO5 VICSA STANDARD; PRT; 124 AA.
AC Q41705;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Early nodulin 5 precursor (N-5) (Fragment).
GN ENOD5.
OS Vicia sativa (Spring vetch) (Tare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Ericoidae; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.
OX NCBI_TaxID=3908;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nigra; TISSUE=Root nodules;
RX MEDLINE=96011756; PubMed=7548828;
RA Vijn I., Yang W.-C., Pallsgaard N., Oestergaard Jensen E.,
RA van Kammen A., Bisseling T.;
RT "VENOD5, VENOD12 and VENOD40 expression during Rhizobium-induced
RT nodule formation on Vicia sativa roots.";
RL Plant Mol. Biol. 28:1111-1119 (1995).
CC -1- FUNCTION: INVOLVED IN THE INFECTION PROCESS DURING THE PLANT-
CC RHIZOBIUM INTERACTION.

```



```

DR InterPro: IPR000100; Ribonuclease_P.
DR Pfam: PF00825; Ribonuclease_P; 1.
DR ProDom: PD003629; Ribonuclease_P; 1.
DR TIGRFAMs: TIGR00188; rnpA; 1.
DR PROSITE: PS00648; RIBONUCLEASE_P; FALSE NEG.
KW Hydrolase; Nuclease; Endonuclease; RNA processing; RNA-binding;
KW Complete proteome.
SQ SEQUENCE 133 AA; 14481 MW; 214D96C013F8E4FB CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 133;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 KTVVYH 239
DB 27 KTVVYH 32

RESULT 67
H32_TETPY
ID H32_TETPY STANDARD; PRT; 135 AA.
AC P15512;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Histone H3.2.
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RP SEQUENCE.
RX MEDLINE=84289353; PubMed=6432775;
RA Hayashi T., Hayashi H., Fusauchi Y., Iwai K.;
RT "Tetrahymena histone H3. Purification and two variant sequences.";
RL J. Biochem. 95:1741-1749(1984).
CC -1- FUNCTION: Histone H3, along with histone H4, plays a central role
in nucleosome formation.
CC -1- SUBUNIT: The nucleosome is an octamer containing two molecules
each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
bp of DNA.
CC -1- SIMILARITY: Belongs to the histone H3 family.
DR InterPro: IPR004822; Histone_core.
DR InterPro: IPR000164; Histone_H3.
DR Pfam: PF00125; histone; 1.
DR PRINTS: PR00622; HISTONEH3.
DR SMART: SM00428; H3; 1.
DR PROSITE: PS00322; HISTONE_H3_1; 1.
DR PROSITE: PS00959; HISTONE_H3_2; 1.
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;
KW Multigene family.
FT INIT MET
SQ SEQUENCE 135 AA; 15388 MW; 45235F3F3915595C CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 135;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 SAPVSG 63
DB 28 SAPVSG 33

RESULT 68
H33_TETTH
ID H33_TETTH STANDARD; PRT; 135 AA.
AC P41563;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Histone.H3.3 (HV2).
HHT3.

```

```

OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9416724; PubMed=8121802;
RA Thacker T.H., Macgaffey J., Bowen J., Horowitz S., Shapiro D.L.,
RA Gorovsky M.A.;
RT "Independent evolutionary origin of histone H3.3-like variants of
RT animals and Tetrahymena.";
RL Nucleic Acids Res. 22:180-186(1994).
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
IN NUCLEOSOME FORMATION. THIS IS A MACRONUCLEAR REPLACEMENT
VARIANT.
CC -1- SUBUNIT: The nucleosome is an octamer containing two molecules
each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
bp of DNA.
CC -1- SIMILARITY: Belongs to the histone H3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M87305; AAC37188.1; -.
DR PIR: S41501; S41501.
DR InterPro: IPR004822; Histone_core.
DR InterPro: IPR000164; Histone_H3.
DR Pfam: PF00125; histone; 1.
DR PRINTS: PR00622; HISTONEH3.
DR SMART: SM00428; H3; 1.
DR PROSITE: PS00322; HISTONE_H3_1; 1.
DR PROSITE: PS00959; HISTONE_H3_2; 1.
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;
KW Multigene family.
FT INIT MET
SQ SEQUENCE 135 AA; 15352 MW; 98235D1B7C155948 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 135;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 SAPVSG 63
DB 28 SAPVSG 33

RESULT 69
NOS_PEA
ID NOS_PEA STANDARD; PRT; 135 AA.
AC P25226;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Early nodulin 5 precursor (N-5).
GN ENOD5.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Sparkle; TISSUE=Root nodules;
RX MEDLINE=93005665; PubMed=2152123;
RA Scheres B., van Engelen F., van der Knaap E., van de Wiel C.,
RA van Kammen A., Bisseling T.;
RT "Sequential induction of nodulin gene expression in the developing
RT pea nodule.";

```

```

RL Plant Cell 2:687-700(1990).
CC - FUNCTION: INVOLVED IN THE INFECTION PROCESS DURING THE PLANT-
CC RHIZOBIA INTERACTION.
CC - TISSUE SPECIFICITY: INVASION ZONE AND EARLY SYMBIOTIC ZONE.
CC - DEVELOPMENTAL STAGE: EXPRESSED IN THE SECOND STAGE OF ROOT NODULE
CC FORMATION.
CC - SIMILARITY: Contains 1 plastocyanin-like domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; S45139; AAB23536.1; -.
DR PIR; JQ1084; JQ1084.
KW Nodulation; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 135 EARLY NODULIN 5.
FT DOMAIN 24 ? PLASTOCYANIN-LIKE.
FT DOMAIN 88 107 PRO-RICH.
SQ SEQUENCE 135 AA; 15063 MW; C6DC133D51FA32B CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 135;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 ITPLPLS 221
Db 95 ITPLPLS 100

RESULT 70
YPO9 DEIRA STANDARD; PRT; 136 AA.
AC Q9RRT2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein DR2509.
GN DR2509.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 92779;
RA MEDLINE=20036896; PubMed=10567266;
RA White O., Bisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamthyan J.J., Lam P., McDonald L., Utecherback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RT Science 286:1571-1577(1999).
RL - SIMILARITY: BELONGS TO THE UPF0081 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE002080; AAF12050.1; -.
DR PIR; D75265; D75265.

```

```

DR TIGR; DR2509; -.
DR InterPro; IPR005227; Cons_hypoth250.
DR InterPro; IPR006641; YqgFc.
DR Pfam; PF03652; UPF0081; 1.
DR SMART; SM00732; YqgFc; 1.
DR TIGRFAMs; TIGR00250; TIGR00250; 1.
KW Hypothetical protein, Complete proteome.
SQ SEQUENCE 136 AA; 14819 MW; 7595D535302F21DE CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 136;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 EAAAVQ 147
Db 123 EAAAVQ 128

RESULT 71
YQGF HAELN STANDARD; PRT; 139 AA.
AC P43981;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein H10305.
GN H10305.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kellavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utecherback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:486-512(1995).
RL - SIMILARITY: BELONGS TO THE UPF0081 FAMILY. STRONG, TO E. COLI YQGF
AND B. APHIDICOLA (SUBSP. ACYRTHOSIPHON PISUM) BU548.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U32716; AAC21970.1; -.
DR PIR; H64005; H64005.
DR TIGR; H10305; -.
DR InterPro; IPR005227; Cons_hypoth250.
DR InterPro; IPR006641; YqgFc.
DR Pfam; PF03652; UPF0081; 1.
DR SMART; SM00732; YqgFc; 1.
DR TIGRFAMs; TIGR00250; TIGR00250; 1.
KW Hypothetical protein, Complete proteome.
SQ SEQUENCE 139 AA; 15339 MW; A875112A732AC6B4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 139;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      107 QALPAF 112
      |||||
Db      28 QALPAF 33

RESULT 72
HB2A2 TRICR STANDARD; PRT; 141 AA.
ID HB2A2 TRICR
AC P10784;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemoglobin alpha-2 chain (Minor).
OS Triturus cristatus (Great crested newt) (Marty newt).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
OC Triturus.
OX NCBI_TaxID=8323;
RN [1]
RP SEQUENCE.
RX MEDLINE=69207117; PubMed=3242554;
RA Kleinschmidt T., Sgourou J.G., Braunitzer G.;
RT "The first sequenced normal hemoglobin lacking histidine in position
RT 146 of the beta-chains. The primary structures of the major and minor
RT hemoglobin components of the great crested newt (Triturus cristatus,
RT Urodela, Amphibia).";
RL Biol. Chem. Hoppe-Seyler 369:1343-1360(1988).
CC -1- FUNCTION: Involved in oxygen transport from the lung to the
CC various peripheral tissues.
CC -1- SUBUNIT: MINOR HEMOGLOBIN IS AN HETEROTETRAMER OF TWO ALPHA-2
CC CHAINS AND TWO BETA-2 CHAINS.
CC -1- TISSUE SPECIFICITY: Red blood cells.
CC -1- SIMILARITY: Belongs to the globin family.
DR PIR: S02026; S02026.
DR HSSP; P01922; 1820.
DR InterPro; IPR002338; Alpha haem.
DR InterPro; IPR000971; Globin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00612; ALPHAHAE.
DR PROSITE; PS01033; GLOBIN; 1.
KM Heme; Oxygen transport; Transport; Erythrocyte.
FT METAL 58 58 IRON (HEME DISTAL LIGAND).
FT METAL 87 87 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 141 AA; 15714 MW; 4059AC571F483ED6 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      197 GKDUKE 202
      |||||
Db      45 GKDUKE 50

RESULT 73
RIB1 PHOLE STANDARD; PRT; 144 AA.
ID RIB1 PHOLE
AC Q01954;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
DE (lumazine synthase) (Riboflavin synthase beta chain) (Fragment).
GN RIBH.
OS Photobacterium leiognathi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OX NCBI_TaxID=658;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25521; PubMed=1339274;
RX MEDLINE=92360014; PubMed=1339274;
RA Lee C.Y., Meighen E.A.;

```

```

RT      "The lux genes in Photobacterium leiognathi are closely linked with
RT genes corresponding in sequence to riboflavin synthase genes.";
RL Biochem. Biophys. Res. Commun. 186:690-697(1992).
CC -1- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
CC 2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-butanone-4-
CC phosphate yielding 6,7-dimethyl-8-lumazine.
CC -1- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -1- PATHWAY: Riboflavin biosynthesis; last step.
CC -1- SIMILARITY: Belongs to the DMRL synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M90094; AAA73230.1; -.
DR PIR; PC1110; PC1110.
DR HSSP; P11998; 1RNV.
DR HAMAP; MF 00178; -; 1.
DR InterPro; IPR002180; DMRL synthase.
DR Pfam; PF00885; DMRL synthase; 1.
DR ProDom; PD003664; DMRL synthase; 1.
DR TIGRFAMs; TIGR00114; ribH; 1.
KM Riboflavin biosynthesis; Transferase.
FT NON TER 144 144
SQ SEQUENCE 144 AA; 15180 MW; 4A4E820CD960359 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      181 ITAIGA 186
      |||||
Db      75 ITAIGA 80

RESULT 74
DSBH OCEIH STANDARD; PRT; 145 AA.
ID DSBH OCEIH
AC Q08R13;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative protein-disulfide oxidoreductase.
GN OBI163.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=1225376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
CC -1- FUNCTION: Required for disulfide bond formation in some proteins
CC (potential).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: Belongs to the dsbB family. DsbH subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP004597; BAC13119.1; -.
DR HAMAP; MF 00287; -.
DR Pfam; PF02600; Ddbb; 1.
KW Hypothetical protein; Electron transport; Oxidoreductase; Chaperone;
KW Redox-active center; Transmembrane; Complete proteome.
FT TRANSMEM 9
FT TRANSMEM 28 POTENTIAL.
FT TRANSMEM 43 62 POTENTIAL.
FT TRANSMEM 69 86 POTENTIAL.
FT TRANSMEM 115 137 POTENTIAL.
FT DISULFID 38 41 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 145 AA; 16263 MW; 6C3DDFA536A72A35 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 WVOAFL 10
DB 16 WVOAFL 21

RESULT 75
MBL1_ECOLI STANDARD; PRT; 148 AA.
ID MBL1_ECOLI
AC P08097;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Mobilization protein MOB (Protein C).
GN MOB OR C.
OS Escherichia coli.
OG Plasmid Clo DF13.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87248075; PubMed=3596243;
RA van Puten A.J., Jochems G.O., de Lang R., Nijkamp H.J.J.;
RT "Structure and nucleotide sequence of the region encoding the
RT mobilization proteins of plasmid C10DF13."
RL Gene 51:171-178 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86314306; PubMed=3749334;
RA Nijkamp H.O.J., de Lang R., Stultje A.R., van den Elsen P.J.M.,
RA Veltkamp E., van Puten A.J.;
RT "The complete nucleotide sequence of the bacteriocinogenic plasmid
RT C10DF13."
RL Plasmid 16:135-160 (1986).
CC -!- FUNCTION: THIS PROTEIN IS ESSENTIAL TO PROMOTE THE SPECIFIC
CC TRANSFER OF THE PLASMID IN THE PRESENCE OF CONJUGATIVE PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04466; CAA28151.1; -.
DR PIR; B29050; MZEC6.
KM Mobility Protein; Plasmid; Conjugation.
SQ SEQUENCE 148 AA; 15933 MW; 2AA49B9628586989 CRC64;

Query Match 2.3%; Score 6; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 144 AAVQPV 149
DB 52 AAVQPV 57

Search completed: November 25, 2003, 13:40:13
Job time : 21 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:37:34 ; Search time 40 Seconds

(without alignments)  
1696.696 Million cell updates/sec

Title: US-10-057-951-2

Sequence: 1 MLAAVQAFLVSNMLAEAY.....PVDPEGSTPLMGAGTPGA 263

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

SPTREMBL.23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	263	100.0	263	4	096FE7 homo sapien
2	250	95.1	263	4	000318 homo sapien
3	101	38.4	263	4	08NCJ9 Q96CJ9 homo sapien
4	9	3.4	331	16	Q99RQ5 Q99RQ5 staphylococ
5	9	3.4	344	16	Q8NV36 Q8NV36 staphylococ
6	8	3.0	155	16	Q8G5B4 Q8G5B4 bifidobacte
7	8	3.0	475	17	Q972D9 Q972D9 sulfolobus
8	8	3.0	602	16	Q97EB2 Q97EB2 clostridium
9	8	3.0	716	11	P70521 P70521 ratcus norv
10	8	3.0	716	11	Q91XG8 Q91XG8 mus musculi
11	8	3.0	764	16	Q8PFY5 Q8PFY5 xanthomonas
12	8	3.0	2585	5	Q23587 Q23587 caenorhabdi
13	7	2.7	37	5	Q8WT10 Q8WT10 plasmodium
14	7	2.7	37	5	Q8WT10 Q8WT10 plasmodium
15	7	2.7	38	5	Q8TWS9 Q8TWS9 lamellidrac
16	7	2.7	107	5	Q8IK21 Q8IK21 plasmodium

17	7	2.7	112	17	Q8TK63 Q8TK63 methanosarc
18	7	2.7	113	17	Q8PM8 Q8PM8 methanosarc
19	7	2.7	116	5	Q8IFK3 Q8IFK3 rlfia pach
20	7	2.7	120	3	Q13544 Q13544 saccharomyc
21	7	2.7	129	13	Q42428 Q42428 lates calca
22	7	2.7	135	10	Q94DN5 Q94DN5 oryza sativ
23	7	2.7	139	17	Q8TIC3 Q8TIC3 methanosarc
24	7	2.7	144	16	Q55181 Q55181 synecocyst
25	7	2.7	147	8	Q94XU2 Q94XU2 crotalus ve
26	7	2.7	147	8	Q94XU1 Q94XU1 crotalus du
27	7	2.7	147	8	Q94XU3 Q94XU3 crotalus un
28	7	2.7	150	16	Q8ZK32 Q8ZK32 salmonella
29	7	2.7	150	16	Q8Z124 Q8Z124 salmonella
30	7	2.7	151	2	Q8KUE6 Q8KUE6 corynebacte
31	7	2.7	154	16	Q8NTP9 Q8NTP9 pyrobaculum
32	7	2.7	158	17	Q8Z2E3 Q8Z2E3 arabidopsis
33	7	2.7	164	10	Q9C692 Q9C692 deinococcus
34	7	2.7	179	16	Q9RZD3 Q9RZD3 alcaligenes
35	7	2.7	192	2	Q44013 Q44013 rhizobium 1
36	7	2.7	197	16	Q98M02 Q98M02 propionibac
37	7	2.7	209	2	Q931G1 Q931G1 bacillus su
38	7	2.7	222	16	Q9RT90 Q9RT90 schizosacch
39	7	2.7	226	16	P96683 P96683 bacillus su
40	7	2.7	229	3	Q9UB2 Q9UB2 clostridium
41	7	2.7	266	11	Q8BH49 Q8BH49 mus musculi
42	7	2.7	272	16	Q8XNE0 Q8XNE0 clostridium
43	7	2.7	273	16	Q9A8M0 Q9A8M0 clostridium
44	7	2.7	282	5	Q9VQU4 Q9VQU4 dirosophila
45	7	2.7	285	5	Q9N3D7 Q9N3D7 caenorhabdi
46	7	2.7	286	16	Q99WRO Q99WRO staphylococ
47	7	2.7	286	16	Q8NYC6 Q8NYC6 staphylococ
48	7	2.7	288	16	Q8PQJ9 Q8PQJ9 xanthomonas
49	7	2.7	288	16	Q8PEA4 Q8PEA4 xanthomonas
50	7	2.7	290	16	Q986W0 Q986W0 rhizobium 1
51	7	2.7	291	17	Q97Y08 Q97Y08 sulfolobus
52	7	2.7	302	5	Q9VPE5 Q9VPE5 dirosophila
53	7	2.7	303	5	Q23364 Q23364 caenorhabdi
54	7	2.7	314	5	Q17035 Q17035 caenorhabdi
55	7	2.7	322	2	Q8VQ72 Q8VQ72 bacillus li
56	7	2.7	323	12	Q9U1B0 Q9U1B0 nelson bay
57	7	2.7	326	16	Q8UTK1 Q8UTK1 agrobacteri
58	7	2.7	332	10	Q941W0 Q941W0 oryza sativ
59	7	2.7	332	16	Q8YWT5 Q8YWT5 anabaena sp
60	7	2.7	346	16	Q8YTX8 Q8YTX8 anabaena sp
61	7	2.7	351	2	Q8XY10 Q8XY10 rhizobium e
62	7	2.7	355	5	Q9GUT1 Q9GUT1 caenorhabdi
63	7	2.7	356	16	Q9Z7B1 Q9Z7B1 chlamydia p
64	7	2.7	356	16	Q9JSA2 Q9JSA2 chlamydia p
65	7	2.7	365	16	Q9Z7A9 Q9Z7A9 chlamydia p
66	7	2.7	365	16	Q9J521 Q9J521 chlamydia p
67	7	2.7	365	16	Q8PDM8 Q8PDM8 xanthomonas
68	7	2.7	370	16	Q8D7W9 Q8D7W9 vibrio vuln
69	7	2.7	378	17	Q9HJF4 Q9HJF4 thermoplas
70	7	2.7	379	10	Q8H482 Q8H482 oryza sativ
71	7	2.7	393	4	Q9BRB6 Q9BRB6 homo sapien
72	7	2.7	394	16	Q8PKX1 Q8PKX1 xanthomonas
73	7	2.7	396	2	Q9EVD2 Q9EVD2 neisseria s
74	7	2.7	398	2	Q9XC57 Q9XC57 pseudomonas
75	7	2.7	414	16	Q8CMV9 Q8CMV9 staphylococ
76	7	2.7	421	16	Q8DF75 Q8DF75 vibrio vuln
77	7	2.7	426	10	Q8LQW3 Q8LQW3 oryza sativ
78	7	2.7	429	16	Q8FZ20 Q8FZ20 brucella su
79	7	2.7	430	16	Q98M05 Q98M05 rhizobium 1
80	7	2.7	448	10	Q9S7M4 Q9S7M4 arabidopsis
81	7	2.7	448	16	Q8UPR7 Q8UPR7 agrobacteri
82	7	2.7	449	16	Q9Z0K1 Q9Z0K1 rhizobium m
83	7	2.7	451	4	Q8N6V2 Q8N6V2 homo sapien
84	7	2.7	451	16	Q8R7E2 Q8R7E2 thermosarc
85	7	2.7	453	17	Q8TQ98 Q8TQ98 xanthomonas
86	7	2.7	454	16	Q8PK13 Q8PK13 xanthomonas
87	7	2.7	454	16	Q8P812 Q8P812 xanthomonas
88	7	2.7	463	16	Q8ZND3 Q8ZND3 salmonella
89	7	2.7	463	16	Q8Z517 Q8Z517 salmonella

```

90 7 2.7 465 16 08PMNS 08pmns xanthomonas
91 7 2.7 469 2 093TX8 093tx8 burkholderi
92 7 2.7 483 5 023737 023737 chelonius sp
93 7 2.7 485 10 041832 041832 zea mays (m
94 7 2.7 485 16 08BQ08 08bq08 oceanobacil
95 7 2.7 511 16 08ZBW6 08zbw6 yersinia pe
96 7 2.7 512 16 08ZMK8 08zmk8 salmonella
97 7 2.7 512 16 08ZAD9 08zad9 salmonella
98 7 2.7 515 16 08FEQ3 08fed3 escherichia
99 7 2.7 520 16 08YIU1 08yiu1 bruceella me
100 7 2.7 534 15 09YKAO 09yka0 murine leuk

```

## ALIGNMENTS

```

RESULT 1
096FE7 PRELIMINARY; PRT; 263 AA.
ID 096FE7;
AC 096FE7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Brain;
RC Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; BC011049; AAH11049.1; -.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE; 1.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
KW Hypothetical protein; Glycoprotein; Kringle.
SQ
SEQUENCE 263 AA; 28234 MW; 197C3EE888FA242 CRC64;

```

```

Query Match          100.0%; Score 263; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.2e-269;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MLAAVQAFVLSNMMLAERAYSGGCFWMDNGHLYREDQTSPPAGLRCLNWLDAQSLASAP 60
DB 1 MLAAVQAFVLSNMMLAERAYSGGCFWMDNGHLYREDQTSPPAGLRCLNWLDAQSLASAP 60
QY 61 VSGAGNHSYCRNPEDDPGPMCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOBAS 120
DB 61 VSGAGNHSYCRNPEDDPGPMCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOBAS 120
QY 121 EGPAGDEVQVAPANALPARSEAAVQPIGISOVRNNSKEKDLGLTGLGYLGITMMVI 180
DB 121 EGPAGDEVQVAPANALPARSEAAVQPIGISOVRNNSKEKDLGLTGLGYLGITMMVI 180
QY 181 IIAIAGIILGYSYRKGDKLEQHDQKVCEREMQRIITPLSAFTNPTEIIDEKTVVHT 240
DB 181 IIAIAGIILGYSYRKGDKLEQHDQKVCEREMQRIITPLSAFTNPTEIIDEKTVVHT 240
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263

```

```

RESULT 2
000318 PRELIMINARY; PRT; 263 AA.
ID 000318

```

```

AC 000318;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE WUSGC:DJ515N1.2 protein.
GN WUSGC:DJ515N1.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Du Z., Scheet P., Harper M.;
RT "The sequence of H. sapiens PAC clone RP3-515N1.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AC002073; AAB54054.1; -.
DR HSSP; P00749; 1KDU.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE; FALSE-NEG.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
KW Glycoprotein; Kringle.
SQ
SEQUENCE 263 AA; 28248 MW; 197C3EE88854A242 CRC64;

```

```

Query Match          95.1%; Score 250; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 6.9e-256;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MLAAVQAFVLSNMMLAERAYSGGCFWMDNGHLYREDQTSPPAGLRCLNWLDAQSLASAP 60
DB 1 MLAAVQAFVLSNMMLAERAYSGGCFWMDNGHLYREDQTSPPAGLRCLNWLDAQSLASAP 60
QY 61 VSGAGNHSYCRNPEDDPGPMCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOBAS 120
DB 61 VSGAGNHSYCRNPEDDPGPMCVYSGEAGVPEKPCEDLRCPETTSQALPAFTTEIOBAS 120
QY 121 EGPAGDEVQVAPANALPARSEAAVQPIGISOVRNNSKEKDLGLTGLGYLGITMMVI 180
DB 121 EGPAGDEVQVAPANALPARSEAAVQPIGISOVRNNSKEKDLGLTGLGYLGITMMVI 180
QY 181 IIAIAGIILGYSYRKGDKLEQHDQKVCEREMQRIITPLSAFTNPTEIIDEKTVVHT 240
DB 181 IIAIAGIILGYSYRKGDKLEQHDQKVCEREMQRIITPLSAFTNPTEIIDEKTVVHT 240
QY 241 SQTVPDPOEG 250
DB 241 SQTVPDPOEG 250

```

```

RESULT 3
08NCJ9 PRELIMINARY; PRT; 263 AA.
ID 08NCJ9;
AC 08NCJ9;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ90207.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Mammary Gland;
RC Iisogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,

```

RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,  
 RA Yamamoto J., Makamatsu A., Nakamura Y., Kojima S., Nagahara K.,  
 RA Masuno Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,  
 RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.,  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -i- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
 DR EMBL: AK074688; BAC1140.1; -.  
 DR InterPro: IPR000001; Kringle.  
 DR Pfam: PF00051; Kringle; 1.  
 DR PRINTS: PR00018; KRINGLE.  
 DR ProDom: PD000395; Kringle; 1.  
 DR SMART: SM00130; KR; 1.  
 DR PROSITE: PS00021; KRINGLE\_1; 1.  
 DR PROSITE: PS00070; KRINGLE\_2; 1.  
 KW Hypothetical protein; Glycoprotein; Kringle.  
 SQ SEQUENCE 263 AA; 28104 MW; 73A9294D51426C3 CRC64;

Query Match 38.4%; Score 101; DB 4; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-98;  
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 KKDGLTGLGVGITMMVITIIAIGAGIILGYSYKRGKDKKEHDKVCEEMORITLPLSA 222  
 DB 163 KKDGLTGLGVGITMMVITIIAIGAGIILGYSYKRGKDKKEHDKVCEEMORITLPLSA 222  
 QY 223 FTNPTCEIVDEKTVVHTSQTPVDPQESSTPLMGAGTPGA 263  
 DB 223 FTNPTCEIVDEKTVVHTSQTPVDPQESSTPLMGAGTPGA 263

## RESULT 4

Q99R05 PRELIMINARY; PRT; 331 AA.

ID Q99R05;  
 AC Q99R05;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein SAV2372.  
 GN SAV2372 OR SA2162.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OS Staphylococcus aureus (strain N315).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=158878; 158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama T., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshoyama A.,  
 RA Mizutani Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuuchi J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus."  
 RL Lancet 357:1225-1240(2001).  
 CC -i- CORFACTOR: FAD (BY SIMILARITY).  
 DR EMBL: AP003365; BAB58534.1; -.  
 DR EMBL: AP003137; BAB43464.1; -.  
 DR InterPro: IPR001327; FAD\_pyr\_redox.  
 DR InterPro: IPR000103; Pyridine\_redox\_2.  
 DR Pfam: PF00070; pyr\_redox; 1.  
 DR PRINTS: PR00368; FADPNR.  
 DR PRINTS: PR00469; PNDRDTASEII.  
 KW FAD; Flavoprotein; Oxidoreductase; Hypothetical protein;  
 KW Complete proteome.  
 SQ SEQUENCE 331 AA; 37034 MW; FCBBD21214107FEA CRC64;

Query Match 3.4%; Score 9; DB 16; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 IIAIGAGII 189  
 DB 97 IIAIGAGII 105

## RESULT 5

Q8NV36 PRELIMINARY; PRT; 344 AA.

ID Q8NV36;  
 AC Q8NV36;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE MW2294 protein.  
 GN MW2294.  
 OS Staphylococcus aureus (strain MW2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.,  
 RT "Genome and virulence determinants of high virulence community-  
 acquired MRSA."  
 RL Lancet 359:1819-1827(2002).  
 DR EMBL: AP004830; BAB96159.1; -.  
 DR InterPro: IPR000759; Adnrx\_reductase.  
 DR InterPro: IPR001327; FAD\_pyr\_redox.  
 DR InterPro: IPR000103; Pyridine\_redox\_2.  
 DR InterPro: IPR003042; Kmg\_moxxygenase.  
 DR Pfam: PF00070; pyr\_redox; 1.  
 DR PRINTS: PR00419; ADXRDTASE.  
 DR PRINTS: PR00368; FADPNR.  
 DR PRINTS: PR00469; PNDRDTASEII.  
 DR PRINTS: PR00420; KMGMOXGNASE.  
 KW Complete proteome.  
 SQ SEQUENCE 344 AA; 38174 MW; 26E33154834C3C61 CRC64;

Query Match 3.4%; Score 9; DB 16; Length 344;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 IIAIGAGII 189  
 DB 110 IIAIGAGII 118

## RESULT 6

Q8G5B4 PRELIMINARY; PRT; 155 AA.

ID Q8G5B4;  
 AC Q8G5B4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN B1101.  
 OS Bifidobacterium longum.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;  
 OC Bifidobacteriaceae; Bifidobacterium.  
 OX NCBI\_TaxID=216816;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCC 2705;  
 RX MEDLINE=22294977; PubMed=12381787;  
 RA Scheil M.A., Karamantzou M., Snel B., Vilanova D., Berger B.,  
 RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,  
 RA Pidmore R.D., Arigoni F.,  
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation  
 to the human gastrointestinal tract."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).  
 DR EMBL: AE014733; AAN24909.1; -.

```

KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 155 AA; 17838 MW; 25CC99676ABBBF8 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 155;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 RITPLPLSA 222
   |||||
Db 130 RITPLPLSA 137

RESULT 7
Q972D9 PRELIMINARY; PRT; 475 AA.
ID Q972D9
AC Q972D9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ST1189.
GN ST1189.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / ?;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagisi M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000985; BAB66230.1; -.
KM Hypothetical protein: Complete proteome.
SQ SEQUENCE 475 AA; 54559 MW; 9F08E45ED90374E9 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 17; Length 475;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EIVDEKTV 236
   |||||
Db 67 EIVDEKTV 74

RESULT 8
Q97EB2 PRELIMINARY; PRT; 602 AA.
ID Q97EB2
AC Q97EB2;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ATP-dependent Zn protease, FISH.
GN CAC3202.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Moelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitt J., Wolf Y.I.,
RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing

```

```

RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007816; AAK81138.1; -.
DR MEROPS; M41.009; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR005936; FISH.
DR InterPro; IPR000642; Peptidase_M41.
DR Pfam; PF01434; Peptidase_M41; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfam; TIGR01241; FISH_fam; 1.
DR PROSITE; PS00674; AAA; 1.
KM Protease; Complete proteome.
SQ SEQUENCE 602 AA; 66211 MW; 5E949978B553A50 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 602;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VSGEAGVP 91
   |||||
Db 213 VSGEAGVP 220

RESULT 9
P70521 PRELIMINARY; PRT; 716 AA.
ID P70521
AC P70521;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Macrophage stimulating protein precursor.
GN MSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Liver;
RX MEDLINE=97011126; PubMed=8858136;
RA Ohsuhiro K., Iwama A., Matsuno K., Ezaki T., Sakamoto O., Hamaguchi I.,
RA Takasu N., Suda T.;
RT "Molecular cloning of Rat Macrophage-stimulating protein and its
RT involvement in the Male Reproductive System.";
RL Biochem Biophys Res Commun. 227:273-280 (1996).
CC -1-SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.
DR EMBL; X95096; CAA64473.1; -.
DR HSSP; P00747; 1KRN.
DR MEROPS; S01.975; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003609; Pan_app.
DR InterPro; IPR003966; Prochrombin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00051; kringle; 4.
DR Pfam; PF00024; PAN; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR PRINTS; PR01505; PROTHROMBIN.
DR ProDom; PD000395; Kringle; 4.
DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN_AP; 1.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00070; KRINGLE_2; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KM Glycoprotein; Hydrolase; Kringle; Protease; Serine protease; Signal.
RT SIGNAL 1 31 POTENTIAL.

```

```

SQ SEQUENCE 716 AA; 80733 MW; 06B7DF3EF56D921F CRC64;
Query Match 3.0%; Score 8; DB 11; Length 716;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 DPGPWCY 83
Db 163 DPGPWCY 170

RESULT 10
Q91XG8 PRELIMINARY; PRT; 716 AA.
AC Q91XG8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hepatocyte growth factor-like.
GN HGFL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.
DR EMBL; BC010551; AAH10551.1; -.
DR HSSP; P00761; IAN1.
DR MGD; MGI:96080; Hgfl.
DR InterPro; IPR001314; CHYMOTRYPSIN.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR003609; Pan app.
DR InterPro; IPR003966; PROTHOMBIN.
DR Pfam; PF00051; Kringle; 4.
DR Pfam; PF00024; PAN; 1.
DR Pfam; PF00089; TRYPSIN; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PRINTS; PRO0018; KRINGLE.
DR PRINTS; PRO1505; PROTHROMBIN.
DR PRODOM; PD000395; Kringle; 4.
DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN AP; 1.
DR SMART; SM00020; TRYF SPC; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS50070; KRINGLE_2; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Glycoprotein; Hydrolase; Kringle; Protease; Serine protease.
SQ SEQUENCE 716 AA; 80693 MW; 12474C48A7DB46D CRC64;

Query Match 3.0%; Score 8; DB 11; Length 716;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 DPGPWCY 83
Db 163 DPGPWCY 170

RESULT 11
Q8PFY5 PRELIMINARY; PRT; 764 AA.
AC Q8PFY5;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein XAC3837.
GN XAC3837.

```

```

OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxId=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camavari F., Cardoso J., Chamberg F., Clapina L.P.,
RA Clarel J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Wenck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira L.A.F., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola H.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Senbal J.C., Kitajima J.P.,
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE012033; AAM3679.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
KW Hypothetical protein; Complete Proteome.
SQ SEQUENCE 764 AA; 85179 MW; D393176D1885646 CRC64;

Query Match 3.0%; Score 8; DB 16; Length 764;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 VQVFAPAN 135
Db 568 VQVFAPAN 575

RESULT 12
Q23587 PRELIMINARY; PRT; 2585 AA.
AC Q23587;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein ZK783.1.
GN ZK783.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabdilitida;
OC Rhabdilitidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Faveillo A., Vaudin M.;
RT "The sequence of C. elegans cosmid ZK783.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;

```

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: U13646; AAC24418.2; -.  
DR HSSP: P00736; 1APO.  
DR Wormpep; ZK783.1; CE25695.  
DR InterPro; IPR00152; Asx\_hydroxyl.  
DR InterPro; IPR00181; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR001507; Endoglin/CD105.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00008; EGF\_15.  
DR SMART; SM00181; EGF; 30.  
DR SMART; SM00179; EGF\_Ca; 16.  
DR SMART; SM00241; ZP\_1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 9.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 22.  
DR PROSITE; PS01187; EGF\_Ca; 13.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KM Hypothetical protein; EGF-like domain.  
SQ SEQUENCE 2585 AA; 271205 MW; 5EFD32B769CAC5B CRC64;

Query Match 3.0%; Score 8; DB 5; Length 2585;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 SGEAGVPE 92  
|||  
Db 1269 SGEAGVPE 1276

## RESULT 13

Q8WT10 PRELIMINARY; PRT; 37 AA.  
ID O8WT10;  
AC O8WT10;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE ETRAMP10.1 protein (Fragment).  
GN ETRAMP10.1.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=6329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21036610; PubMed=11163452;  
RA Spielmann T., Beck H.P.;  
RT "Analysis of stage-specific transcription in Plasmodium falciparum reveals a set of genes exclusively transcribe in ring stage parasites.";  
RL Mol. Biochem. Parasitol. 111:453-458(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RA Spielmann T., Beck H.P.;  
RT "Etiramps, a new Plasmodium falciparum gene family coding for highly charged membrane proteins located at the parasite-host cell interface.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF420672; CAD12622.1; -.  
FT NON\_TER 1  
FT NON\_TER 37  
SQ SEQUENCE 37 AA; 4035 MW; F0734FEAC69E4B58 CRC64;

Query Match 2.7%; Score 7; DB 5; Length 37;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 248 QEGSTPL 254  
|||  
Db 26 QEGSTPL 32

## RESULT 14

Q9TWT0 PRELIMINARY; PRT; 37 AA.  
ID Q9TWT0;  
AC Q9TWT0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE 44 kDa hemoglobin A2 chain (Fragment).  
OS Lamellibrachia sp. (Deep-sea giant tube worm).  
OC Eukaryota; Metazoa; Pogonophora; Vestimentifera; Basibranchia;  
OC Lamellibrachida; Lamellibrachidae; Lamellibrachia.  
OX NCBI\_TaxID=6424;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=93320735; PubMed=7763791;  
RA Suzuki T., Takagi T., Ohta S.;  
RL Zool. Sci. 10:141-146(1993).  
DR InterPro; IPR000971; Globin.  
DR Pfam; PF00042; globin; 1.  
KM Heme; Oxygen transport; Transport.  
SQ SEQUENCE 37 AA; 4456 MW; 39ABA4F978633CD0E CRC64;

Query Match 2.7%; Score 7; DB 5; Length 37;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AEAYSGS 23  
|||  
Db 16 AEAYSGS 22

## RESULT 15

Q9TWS9 PRELIMINARY; PRT; 38 AA.  
ID Q9TWS9;  
AC Q9TWS9;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE 44 kDa hemoglobin A2 chain (Fragment).  
OS Lamellibrachia sp. (Deep-sea giant tube worm).  
OC Eukaryota; Metazoa; Pogonophora; Vestimentifera; Basibranchia;  
OC Lamellibrachida; Lamellibrachidae; Lamellibrachia.  
OX NCBI\_TaxID=6424;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=93320735; PubMed=7763791;  
RA Suzuki T., Takagi T., Ohta S.;  
RL Zool. Sci. 10:141-146(1993).  
DR InterPro; IPR000971; Globin.  
DR Pfam; PF00042; globin; 1.  
KM Heme; Oxygen transport; Transport.  
SQ SEQUENCE 38 AA; 4528 MW; 8DBA4ED6A1D21AB8 CRC64;

Query Match 2.7%; Score 7; DB 5; Length 38;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AEAYSGS 23  
|||  
Db 17 AEAYSGS 23

## RESULT 16

Q8IK21 PRELIMINARY; PRT; 107 AA.  
ID Q8IK21;  
AC Q8IK21;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Early transcribed membrane protein.  
GN Pf10\_0019.  
OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriaman M., Hyman R.W.,  
 RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shalom S.J., Sub B., Peterson J., Anguillo S.,  
 RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrall B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 falciparum";  
 RL Nature 419:498-511(2002).  
 DR EMBL: AE014829; AAN35217.1; -;  
 SQ SEQUENCE 107 AA; 11267 MW; A808FBD6C01923A CRC64;

Query Match 2.7%; Score 7; DB 5; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 QGSGTPL 254  
 Db 96 QGSGTPL 102

RESULT 17  
 ID Q8TK63 PRELIMINARY; PRT; 112 AA.  
 AC Q8TK63;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Cobalt ABC transporter, solute-binding protein.  
 GN CBIN OR MA3553.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZA / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,  
 RA Fitzhugh W., Galvo S., Engels R., Smitnov S., Atwood D., Brown A.,  
 RA Allen N., Naylor J., Stange-Thomann N., DeRellano K., Johnson R.,  
 RA Lincon L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuelster H.C., Krzycki J.A.,  
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
 RA Metcalf W.W., Birren B.;  
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 and physiological diversity";  
 RL Genome Res. 12:532-542(2002).  
 DR EMBL: AB011064; AAM06915.1; -;  
 DR InterPro: IPR003705; CBIN.  
 DR Pfam: PF02553; CBIN; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 112 AA; 12244 MW; C98FA33AA9468B55 CRC64;

Query Match 2.7%; Score 7; DB 17; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 AIGAGII 189  
 |||||

Db 75 AIGAGII 81

RESULT 18  
 ID Q8PZM8 PRELIMINARY; PRT; 113 AA.  
 AC Q8PZM8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Cobalt transport protein.  
 GN CBIN OR MM0464.  
 OS Methanosarcina mazei (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RX MEDLINE=22120827; PubMed=12125824;  
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,  
 RA Martinez-Arias R., Henne A., Wietzer A., Baumer S., Jacobi C.,  
 RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,  
 RA Bartschary A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
 RA Fritz H.-J., Gottschalk G.;  
 RT "The genome of Methanosarcina mazei: evidence for lateral gene  
 transfer between Bacteria and Archaea.";  
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL: AE013272; AAM30160.1; -;  
 DR InterPro: IPR003705; CBIN.  
 DR Pfam: PF02553; CBIN; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 113 AA; 12428 MW; 16AE084D3D3114B0 CRC64;

Query Match 2.7%; Score 7; DB 17; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 AIGAGII 189  
 |||||  
 Db 75 AIGAGII 81

RESULT 19  
 ID Q8IFK3 PRELIMINARY; PRT; 116 AA.  
 AC Q8IFK3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hemoglobin A2 chain (fragment).  
 GN HBZ2.  
 OS Riftia pachyptila (Tube worm).  
 OC Eukaryota; Metazoa; Pogonophora; Vestimentifera; Axonobranchia;  
 OC Riftidia; Riftidiidae; Riftia.  
 OX NCBI\_TaxID=6426;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bailey X., Jollivet D., Vanin S., Deutech J., Zal F., Lallier F.,  
 RA Toulmond A.;  
 RT "Evolution of the sulfide-binding function within the globin  
 RT multigene family of the deep-sea hydrothermal vent tubeworm Riftia  
 pachyptila.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ439733; CAD29155.1; -;  
 DR NON\_TER 1  
 FT NON\_TER 1  
 FT 116 116  
 SQ SEQUENCE 116 AA; 13011 MW; 8A45B0DFD8543CB7 CRC64;

Query Match 2.7%; Score 7; DB 5; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      17 AEAAYSG 23
      |||||
Db      2 AEAAYSG 8

RESULT 20
013544      PRELIMINARY;      PRT;      120 AA.
ID 013544;
AC 013544;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE YLR302CP.
GN YLR302C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97313267; Pubmed=9169871;
RA Johnston M., Hillier L., Riles L., Dubois E., Dusterhoft A.,
RA Jones V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
RA Entian K.D., Floeth M., Goffeau A., Heblung U., Henmann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Miesga T., Mostl D.,
RA Louis E.J., Messenguy F., Mewes H.W., Miesga T., Pohl T.M.,
RA Muller-Auer S., Nentwich U., Obermayer B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M.,
RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA Verendeels F., Voet M., Volckaert G., Voss H., Wambut R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohlseil J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Bailey A.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cherry J.M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U17243; AAB67353.1; -.
DR SCD: S0004293; YLR302C.
SQ SEQUENCE 120 AA; 14090 MW; 879002561CTDAC67 CRC64;

Query Match      2.7%; Score 7; DB 3; Length 120;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      229 EIVDEKT 235
      |||||
Db      25 EIVDEKT 31

RESULT 21
042428      PRELIMINARY;      PRT;      129 AA.
ID 042428;
AC 042428;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Tyrosine hydroxylase (Fragment).
GN TH.
OS Lates calcarifer (Barramundi).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Centropomidae; Lates.

OX NCBI_TaxId=8187;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Barramundi;
RA Collet C., Candy J., Sara V.;
RT "Tyrosine hydroxylase and insulin-like growth factor-II but not
RT insulin are adjacent in the teleost species barramundi (Lates
RT calcarifer).";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF007942; AAB64194.1; -.
DR HSSP: P04177; 1TOH.
DR InterPro: IPR001273; Aaa_hydroxylase.
DR Pfam: PF00351; bioplerin_H.1.
DR PRINTS: PR00372; FYMHYKXLAASE.
FT NON_TER
SQ SEQUENCE 129 AA; 14678 MW; 1AE29C7530ABAD78 CRC64;

Query Match      2.7%; Score 7; DB 13; Length 129;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      142 EAAAYQP 148
      |||||
Db      47 EAAAYQP 53

RESULT 22
094DNS      PRELIMINARY;      PRT;      135 AA.
ID 094DNS;
AC 094DNS;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-OCT-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE P0483G10.8 protein (P0401G10.30 protein).
GN P0483G10.8 OR P0401G10.30.
OS Oryza sativa (Rice), and
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxId=4530, 39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RT clone:P0401G10.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RT clone:P0401G10.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003263; BAB63614.1; -.
DR EMBL: AP003268; BAB89022.1; -.
DR Gramene: Q94DNS; -.
SQ SEQUENCE 135 AA; 13916 MW; 4B06F9BA65ABAD95 CRC64;

Query Match      2.7%; Score 7; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      139 ARSEANA 145
      |||||
Db      23 ARSEANA 29

RESULT 23
08TLC3      PRELIMINARY;      PRT;      139 AA.
ID 08TLC3;
AC 08TLC3;
```



```

DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Predicted protein.
GN MA3113.
OS Methanosarcina acetivorans.
OC Archaeae; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OK NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeRellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Titrill A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Mercalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011014; AA06486.1; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 15457 MW; 37D003E9BA0C8754 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 17; Length 139;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 DEKTVV 238
Db 39 DEKTVV 45

RESULT 24
05181
ID 05181 PRELIMINARY; PRT; 144 AA.
AC 05181;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein sir0491.
GN SIR0491.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OK NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 6803;
RA Tabata S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9612752; PubMed=8590279;
RA Kaneo T., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64k to 92k of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hosokawa M., Sugitara M., Sasamoto S., Kimura S.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

```

```

RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D64001; BAA10321.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 144 AA; 16726 MW; B6993A507435742 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 144;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TLGVVLG 174
Db 40 TLGVVLG 46

RESULT 25
094XU2
ID 094XU2 PRELIMINARY; PRT; 147 AA.
AC 094XU2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE NADH dehydrogenase subunit 5 (Fragment).
GN ND5.
OS Crocatalus vegrandis.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crocotalinae; Crocatalus.
OK NCBI_TaxID=425902;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy R.W., Fu J., Lathrop A., Feltham J.V., Kovac V.;
RT "Phylogeny of the rattlesnakes (Crocatalus and Sistrurus) inferred from
RT sequences of five mitochondrial DNA genes.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF259211; AAL25699.1; -.
DR InterPro: IPR001515; Oxidored_q1_N.
DR Pfam; PF00662; oxidored_q1_N; 1.
KW Mitochondrion.
FT NON TER 147 147
SQ SEQUENCE 147 AA; 16924 MW; 925734D1CD0B45 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 8; Length 147;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 ITMMVIT 181
Db 122 ITMMVIT 128

RESULT 26
094XU1
ID 094XU1 PRELIMINARY; PRT; 147 AA.
AC 094XU1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE NADH dehydrogenase subunit 5 (Fragment).
GN ND5.
OS Crocatalus durissus (tropical rattlesnake).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crocotalinae; Crocatalus.
OK NCBI_TaxID=8731;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy R.W., Fu J., Lathrop A., Feltham J.V., Kovac V.;

```

```

RT "Phylogeny of the rattlesnakes (Crotalus and Sistrurus) inferred from
RT sequences of five mitochondrial DNA genes."
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF259212; AAL25700.1; -.
DR InterPro: IPR001516; Oxidored_q1_N.
DR Pfam: PF00662; Oxidored_q1_N.1.
KM Mitochondrion.
FT NON TER 147
SQ SEQUENCE 147 AA; 16924 MW; 925734D12CD0BB45 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 8; Length 147;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 ITMMVII 181
Db 122 ITMMVII 128

RESULT 27
Q94XU3 PRELIMINARY; PRT; 147 AA.
AC Q94XU3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE NADH dehydrogenase subunit 5 (Fragment).
GN ND5.
OS Crotalus unicolor.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosaurs; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Crotalus.
OX NCBI_Taxid=125874;
RN [1]_Taxid=125874;
RP SEQUENCE FROM N.A.
RA Murphy R.W., Fu J., Lathrop A., Feltham J.V., Kovac V.;
RT "Phylogeny of the rattlesnakes (Crotalus and Sistrurus) inferred from
RT sequences of five mitochondrial DNA genes."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF259210; AAL25698.1;
DR InterPro: IPR001516; Oxidored_q1_N.
DR Pfam: PF00662; Oxidored_q1_N.1.
KM Mitochondrion.
FT NON TER 147
SQ SEQUENCE 147 AA; 16781 MW; 716689AB226EBB4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 8; Length 147;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 ITMMVII 181
Db 122 ITMMVII 128

RESULT 28
Q8ZK32 PRELIMINARY; PRT; 150 AA.
ID Q8ZK32;
AC Q8ZK32;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Putative cytoplasmic protein.
GN YOGK OR STM4468.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RX
```

```

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvarey E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
DR EMBL: AE008909; AAL21287.1; -.
DR InterPro: IPR004375; Cons_hypoth22.
DR Pfam: PF04074; DUF386; 1.
DR TIGRFAMs: TIGR00022; TIGR00022; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 150 AA; 16710 MW; 4405D0D69635FE3D CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 150;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 VDEKTVV 237
Db 107 VDEKTVV 113

RESULT 29
Q82124 PRELIMINARY; PRT; 150 AA.
ID Q82124;
AC Q82124;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein STY4806.
GN STY4806.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_Taxid=601;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Kung'u K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
RA Baker S., Basham D., Brooks K., Dillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Chow L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quait M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL: AL627283; CAD06928.1; -.
DR InterPro: IPR004375; Cons_hypoth22.
DR Pfam: PF04074; DUF386; 1.
DR TIGRFAMs: TIGR00022; TIGR00022; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 150 AA; 16638 MW; 4D73836B375E1DED CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 150;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 VDEKTVV 237
Db 107 VDEKTVV 113

RESULT 30
Q8KUE6 PRELIMINARY; PRT; 151 AA.
ID Q8KUE6;
AC Q8KUE6;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
RX
```

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Lrp-like regulator.  
GN Lrp.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
CC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22077265; PubMed=12081967;  
RA Kennerknecht N., Sam H., Yen W.R., Patek M., Sailer M.H. Jr.,  
RA Eggeling L.;  
RT "Export of L-Isoleucine from Corynebacterium glutamicum: a Two-Gene-  
RT Encoded Member of a New Translocator Family.";  
RL J. Bacteriol. 184:3947-3956(2002).  
CC -1- SIMILARITY: BELONGS TO THE ASNC FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL; AF454053; AAM46687.1; -;  
DR InterPro; IPR000485; HTH\_Asnc.  
DR Pfam; PF01037; ASNC\_trans\_reg; 1.  
DR PRINTS; PR00033; HTHASNC.  
DR SMART; SM00344; HTH\_Asnc; 1.  
KM DNA-Binding; Transcription; Transcription regulation.  
SQ SEQUENCE 151 AA; 16702 MW; 76B2136BD31CB8A5 CRC64;

Query Match 2.7%; Score 7; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 GILIGYS 193  
Db 47 GILIGYS 53

RESULT 31  
Q8NTP9 PRELIMINARY; PRT; 154 AA.  
ID Q8NTP9  
AC Q8NTP9;  
DT 01-OCT-2002 (TReMBLrel. 22, Created)  
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Transcriptional regulators.  
GN CG10257.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
CC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
RA Nakagawa S.;  
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032";  
RL Submitted (May-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF005274; BAB97650.1; -;  
DR InterPro; IPR000485; HTH\_Asnc.  
DR Pfam; PF01037; ASNC\_trans\_reg; 1.  
DR PRINTS; PR00033; HTHASNC.  
KM Complete proteome.  
SQ SEQUENCE 154 AA; 17124 MW; 06B603D986DD790E CRC64;

Query Match 2.7%; Score 7; DB 16; Length 154;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 GILIGYS 193  
Db 50 GILIGYS 56

RESULT 32  
Q8Z2E3 PRELIMINARY; PRT; 158 AA.

AC Q8Z2E3;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE PAREP.  
GN PAE0302.  
OS Pyrobaculum aerophilum.  
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
CC Thermoproteaceae; Pyrobaculum.  
OX NCBI\_TaxID=13773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=IM2 / ATCC 51768 / DSM 7523;  
RL PubMed=11792869;  
RA Flitz-Gibson S.T., Ladner H., Kim U.-J., Stettler K.O., Simon M.I.,  
RA Miller J.H.;  
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
RT aerophilum";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
DR EMBL; AE009761; AAL62698.1; -;  
KM Complete proteome.  
SQ SEQUENCE 158 AA; 18145 MW; CF35FE78E609CB08 CRC64;

Query Match 2.7%; Score 7; DB 17; Length 158;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LLAAYG 21  
Db 99 LLAAYG 105

RESULT 33  
Q9C892 PRELIMINARY; PRT; 164 AA.  
ID Q9C892  
AC Q9C892;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE Hypothetical 18.5 kDa protein.  
GN F7A10.6.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eusids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=cv. Columbia;  
RX MEDLINE=21016719; PubMed=11130712;  
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurat D.B., Kwan A., Lam B.,  
RA Langin-Hooper S., Lee A., Lee J.M., Benz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Matzilai A.,  
RA Millscher U., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tambung G., Toriumi M.J., Town C.D.,  
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
RT thaliana";  
RL Nature 408:816-820(2000).  
DR EMBL; AC027034; AAG51579.1; -;  
KM Hypothetical protein.  
SQ SEQUENCE 164 AA; 18498 MW; 97C054B6B82B80BA CRC64;

Query Match 2.7%; Score 7; DB 10; Length 164;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 KKDGLT 169  
 |||||  
 Db 145 KKDGLT 151

RESULT 34  
 Q9RZD3 PRELIMINARY; PRT; 179 AA.  
 AC Q9RZD3;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Cobinamide kinase/cobinamide phosphate guanylyltransferase.  
 GN DRA0020.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 RX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RC MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Nelson M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Uetreck T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1."  
 RL Science 286:1571-1577(1999).  
 DR EMBL; AB01862; AAF1282.1; -.  
 DR HSSP; Q05599; ICGK.  
 DR TIGR; DRA0020; -.  
 DR InterPro: IPR003203; COBU.  
 DR Pfam; PF02283; COBU; 1.  
 DR Transfers: Kinase; Nucleotidyltransferase; Complete proteome.  
 SQ SEQUENCE 179 AA; 19525 MW; 8B37488249C55E41 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VSNMILA 17  
 |||||  
 Db 96 VSNMILA 102

RESULT 35  
 Q44013 PRELIMINARY; PRT; 192 AA.  
 AC Q44013;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Granule-associated protein (PHASIN).  
 GN PHAP.  
 OS Alcaligenes eutrophus (Ralstonia eutropha).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.  
 RX NCBI\_TaxID=510;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wicczorek R., Pries A., Steinbuechel A., Mayer F.,  
 RT "Analysis of a 24 kDa protein associated with the polyhydroxyalkanoic  
 RT acid granules in Alcaligenes eutrophus."  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=H16;  
 RA Hanley S.Z., Pappin D.J.C., White A.J., Elborough K.M., Sibbas A.R.;  
 RT "DNA and protein sequencing indicate that the carboxy-terminal region  
 RT of phasin implicated in polyalkanoic acid granule binding does not  
 RT form part of the in vivo protein."  
 RL FEBS Lett. 0:0-0(1999).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RA York G.M., Stubbe J., Sinskey A.J.;  
 RT "The Ralstonia eutropha phasin Phap promotes synthesis of  
 RT polyhydroxybutyrate throughout the period of cultivation conditions."  
 RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; X85729; CAA59734.1; -.  
 DR EMBL; X85729; CAA59734.1; -.  
 DR EMBL; AF079155; AAC78327.1; -.  
 DR EMBL; AF314206; AAC33635.1; -.  
 DR InterPro: IPR001778; POA\_allergenc.  
 DR PRINTS; PR00833; POAALLERGEN.  
 SQ SEQUENCE 192 AA; 19966 MW; 3A69C094ED9A3ECE CRC64;

Query Match 2.7%; Score 7; DB 2; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 AAAPVPV 149  
 |||||  
 Db 69 AAAPVPV 75

RESULT 36  
 Q98M02 PRELIMINARY; PRT; 197 AA.  
 AC Q98M02;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Hypothetical protein ml10797.  
 GN ML10797.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Mesorhizobium.  
 RX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RC MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti."  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AP002995; BAB48311.1; -.  
 DR InterPro: IPR005586; DUF330.  
 DR Pfam; PF03886; DUF330; 1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 197 AA; 20709 MW; 55F29BD677CC6B74 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ASAPVSG 63  
 |||||  
 Db 162 ASAPVSG 168

RESULT 37  
 Q93IG1 PRELIMINARY; PRT; 209 AA.

AC Q9JIG1; 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical 21.9 kDa protein.  
 OS Propionibacterium freudenreichii shermanii.  
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  
 CC Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.  
 OX NCBI\_TaxID=1752;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Roesserer C.A., Huang K., Scott A.I.;  
 RT "Cobalamin biosynthesis in Propionibacterium freudenreichii  
 (shermanii): Isolation and characterization of 16 vitamin B12 genes."  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY033236; AAK67498.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 209 AA; 21906 MW; D1653DB38370DD98 CRC64;

Query Match 2.7%; Score 7; DB 2; Length 209;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TTSQLP 110  
 |||||  
 DB 123 TTSQLP 129

## RESULT 38

Q9RT90

ID Q9RT90 PRELIMINARY; PRT; 222 AA.

AC Q9RT90;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Hypothetical protein DR1875.

GN DR1875

OS Deinococcus radiodurans.

OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;

CC Deinococcaceae; Deinococcus.

OX NCBI\_TaxID=1299;  
 [1]

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RI;

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
 Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 Ketchum C.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 Fraser C.M.;RA "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1.";

RL Science 286:1571-1577 (1999).

DR EMBL; AE002027; AAF11429.1; -

DR TIGR; DR1875; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 222 AA; 22884 MW; 8E542E41B016C09 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 222;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 LASAPVS 62  
 |||||  
 DB 2 LASAPVS 8

RESULT 39  
 P96683 PRELIMINARY; PRT; 226 AA.  
 AC P96683;

DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE YDFP protein.  
 OS Bacillus subtilis.  
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=168;  
 RX MEDLINE=98000887; PubMed=941680;  
 RA Beloin C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,  
 Kaashara Y., Alonso J.C., Le Hegarat F.;  
 RT "Characterization of an lrp-like (ltpc) gene from Bacillus subtilis."  
 RL Mol. Gen. Genet. 256:63-71 (1997).  
 [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Berteiro M.G., Bessieres P., Bolydin A., Borchert S.,  
 Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Choi S.K., Codann J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Eyring J., Fabre C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haech U., Hatwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kaashara Y., Kleier-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Pardo V., Pohl T.M., Portetlelle D., Portwollisch P., Prescott A.M.,  
 Presceman E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche B., Roche M., Sadie Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Yamakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 Viari A., Wambert R., Wedler E., Wedler H., Weitzengger T.,  
 Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis.";

RL

Db 99 SKEKDL 105

## RESULT 40

Q9UBB2 PRELIMINARY; PRT; 229 AA.  
AC Q9UBB2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical 26.7 kDa protein.  
GN SPBC409.12C.  
OS Schizosaccharomyces pombe (Fission Yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972b-;  
RA Lyne M.H., Rajandream M.A., Barrell B.G., Chillingworth T.,  
RA Churcher C.M., 1999) to the EMBL/GenBank/DBJ databases.  
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL109822; CAB52614.1; -;  
DR GeneDB; Spombe; SPBC409.12c; -;  
KW Hypothetical protein.  
SQ SEQUENCE 229 AA; 26714 MW; 937C8D170A55A063 CRC64;

Query Match 2.7%; Score 7; DB 3; Length 229;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 ITPLPSA 222  
Db 192 ITPLPSA 198

## RESULT 41

Q8BH49 PRELIMINARY; PRT; 266 AA.  
AC Q8BH49;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical pleckstrin homology.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cecum, and Hypothalamus;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
DR EMBL; AK033618; BAC28394.1; -;  
DR EMBL; AK039192; BAC30272.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 266 AA; 29218 MW; 4D0C3B71A5D23F4F CRC64;

Query Match 2.7%; Score 7; DB 11; Length 266;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 PANALPA 139  
Db 135 PANALPA 141

## RESULT 42

Q8XNE0 PRELIMINARY; PRT; 272 AA.  
AC Q8XNE0;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Purine nucleoside phosphorylase.  
GN DEOD OR CPE0398.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / Type A;  
RX PubMed=11792842;  
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
RT flesh-eater.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).  
DR EMBL; AP001186; BAB80104.1.  
DR InterPro; IPR001369; Mcap\_FNP.  
DR Pfam; PF00896; Mcap\_FNP; 1.  
KW Complete proteome.  
SQ SEQUENCE 272 AA; 30064 MW; F54F42F21AF6695F CRC64;

Query Match 2.7%; Score 7; DB 16; Length 272;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 MORITLP 219  
Db 90 MORITLP 96

## RESULT 43

Q9A8M0 PRELIMINARY; PRT; 273 AA.  
AC Q9A8M0;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Phage SP01 DNA polymerase-related protein.  
GN C01333.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;  
OC Caulobacteraceae; Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Ulteback T., Tran K., Wolf A., Yatchew J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).  
DR EMBL; AE005808; AK23314.1; -;  
DR TIGR; C01333; -;  
DR InterPro; IPR005273; SP01polnrel.  
DR InterPro; IPR005122; UDNA\_glycylsef.  
DR Pfam; PF03167; UDG; 1.  
DR TIGRPFAM; TIGR00758; SP01polnrel; 1.  
KW Complete proteome.  
SQ SEQUENCE 273 AA; 29214 MW; CE54307D81568194 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 273;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 121 EGPAGE 127  
Db 123 EGPAGE 129

## RESULT 44

Q9VQ04 PRELIMINARY; PRT; 282 AA.  
AC Q9VQ04; 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)  
DE CG3410 protein.  
GN LECTIN-24A OR CG3410.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Mills G.L.G.,  
RA Aroll J.F., Agbayani A., An H.-D., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Broksrein P., Brotier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Paoli B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glied A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,  
RA Jajall M., Kalush F., Kappen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Lau X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Slater E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Stretcher R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach U.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "the genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RU EMBL; AB003579; AAF51070.1;  
DR FLYbase; FBgn040104; lectin-24A.  
DR InterPro; IPR01304; lectin\_C.  
DR Pfam; PF00059; lectin\_c; 1.  
DR SMART; SM00034; CLECT\_1.  
DR PROSITE; PSS0041; C TYPE LECTIN 2; 1.  
SQ SEQUENCE 282 AA; 32320 MW; C0D933D1B54791 CRC64;

Query Match 2.7%; Score 7; DB 5; Length 282;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 47 LNWDAQ 53  
Db 178 LNWDAQ 184

## RESULT 45

Q9N3D7 PRELIMINARY; PRT; 285 AA.  
AC Q9N3D7; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
DE Y54E10BL.2 protein.  
GN Y54E10BL.2  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabdilitidae;  
OC Rhabdilitidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99059613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC024811; AAF60775.1;  
DR WormPep; Y54E10BL.2; CE25434.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR002486; Col\_cuticle\_N.  
DR Pfam; PF01391; Col\_cuticle\_N.  
DR Pfam; PF01484; Col\_cuticle\_N; 1.  
SQ SEQUENCE 285 AA; 27871 MW; 5568D837E02B3815 CRC64;

Query Match 2.7%; Score 7; DB 5; Length 285;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 256 GQAGPG 262  
Db 256 GQAGPG 262

Query Match 2.7%; Score 7; DB 5; Length 282;

RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiratake K.,  
 RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*  
 RT aureus.";  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL; AP003359; BAB56478.1;  
 DR EMBL; AP003130; BAB41529.1;  
 DR InterPro; IPR000600; ROK\_family.  
 DR Pfam; PF00480; ROK; 1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 286 AA; 31683 MW; 6BB4CBAAT7291209 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 QPVIGIS 153  
 Db 59 QPVIGIS 65

RESULT 47  
 Q8NYC6 PRELIMINARY; PRT; 286 AA.  
 AC Q8NYC6; DB 16; Length 286;  
 DT 01-OCT-2002 (TEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)  
 DE 01-OCT-2002 (TEMBLrel. 22, Last annotation update)  
 DE MM0293 protein.  
 GN MM0293.  
 OS *Staphylococcus aureus* (strain MW2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiratake K.,  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 359:1819-1827(2002).  
 DR EMBL; AP004823; BAB94158.1;  
 DR InterPro; IPR000600; ROK\_family.  
 DR Pfam; PF00480; ROK; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 286 AA; 31683 MW; 7291CAF61C497862 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 QPVIGIS 153  
 Db 59 QPVIGIS 65

RESULT 48  
 Q8PJ09 PRELIMINARY; PRT; 288 AA.  
 AC Q8PJ09; DB 16; Length 288;  
 DT 01-OCT-2002 (TEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Polyamine transport protein.  
 GN POTR OR XAC2471.  
 OS *Xanthomonas axonopodis* (pv. citri).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;

RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Camavan F., Cardoso J., Chambergo F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two *Xanthomonas* pathogens with differing  
 RT host specificities.";  
 RL Nature 417:459-463(2002).  
 DR EMBL; AE011884; AAM37322.1;  
 DR InterPro; IPR000515; BPD\_transp.  
 DR Pfam; PF00528; BPD\_transp. 1.  
 DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBER. 1.  
 KM Complete proteome.  
 SQ SEQUENCE 288 AA; 31804 MW; 7716C72128B94B5 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 288;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 215 RITPLPS 221  
 Db 201 RITPLPS 207

RESULT 49  
 Q8PB44 PRELIMINARY; PRT; 288 AA.  
 AC Q8PB44; DB 16; Length 288;  
 DT 01-OCT-2002 (TEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Polyamine transport protein.  
 GN POTR OR XCC2339.  
 OS *Xanthomonas campestris* (pv. campestris).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPPB 528;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Camavan F., Cardoso J., Chambergo F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two *Xanthomonas* pathogens with differing  
 RT host specificities.";  
 RL Nature 417:459-463(2002).  
 DR EMBL; AE012341; AAM41617.1;  
 DR InterPro; IPR000515; BPD\_transp.  
 DR Pfam; PF00528; BPD\_transp. 1.



```

DR PROSITE; PS00402; BPD_TRANS_P_NN_MEMBER; 1.
KW Complete proteome.
SQ SEQUENCE 288 AA; 31877 MW; 1F3D3866C71595A CRC64;

Query Match
Best Local Similarity 2.7%; Score 7; DB 16; Length 288;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 RITLPLS 221
   |||||
   201 RITLPLS 207

RESULT 50
Q986W0 PRELIMINARY; PRT; 290 AA.
AC Q986W0;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE ABC transporter, permease protein.
GN MLT186.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC NCB1_TaxID=381;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Wochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003011; BAB5343.1; -.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF00528; BPD_transp; 1.
DR PROSITE; PS00402; BPD_TRANS_P_NN_MEMBER; 1.
KW Complete proteome.
SQ SEQUENCE 290 AA; 31559 MW; D21F51F73ED7D2 CRC64;

Query Match
Best Local Similarity 2.7%; Score 7; DB 16; Length 290;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 RITLPLS 221
   |||||
   205 RITLPLS 211

RESULT 51
Q97V08 PRELIMINARY; PRT; 291 AA.
AC Q97V08;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein SS02829.
GN SS02829.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCB1_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyer M.J., Chan-Weiner C.C.-Y., Clausen I.G., Curtis B.A.,

```

```

RA De Moore A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet R., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006877; AKA42939.1; -.
DR InterPro; IPR006685; MSion channel.
DR Pfam; PF00924; MS channel; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 291 AA; 31912 MW; 2D1F8A1C3D42820 CRC64;

Query Match
Best Local Similarity 2.7%; Score 7; DB 17; Length 291;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 VITIAIG 185
   |||||
   50 VITIAIG 56

RESULT 52
Q9VVF5 PRELIMINARY; PRT; 302 AA.
AC Q9VVF5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE CG13727 protein.
GN CG13727.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCB1_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoeklin R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrara S., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jakmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liang P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mikhlov G., Milyshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard D., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spadling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

```

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AB003524; AAF49356.1; -;  
 DR FlyBase: FBgn0036711; CG13727.  
 SQ SEQUENCE 302 AA; 31787 MW; 5D373F9C24329B31 CRC64;

Query Match 2.7%; Score 7; DB 5; Length 302;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 SGEAGVP 91  
 Db 176 SGEAGVP 182

## RESULT 53

ID Q23364 PRELIMINARY; PRT; 303 AA.  
 AC Q23364;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical 29.5 kDa protein (Collagen).  
 GN ZC513.8 OR COL-43.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;

RA "SEQUENCE FROM N.A."  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode *C. elegans*: a platform for  
 RT investigating biology. The *C. elegans* Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]

RA "SEQUENCE FROM N.A."  
 RC STRAIN=Bristol N2;  
 RA Wu X., Le T.T.;  
 RT "The sequence of *C. elegans* cosmid ZC513."  
 RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RA "SEQUENCE FROM N.A."  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]

RA Bando T., Ikeda T., Kagawa H.;  
 RT "HOMOTIC PROTEINS, MAB-18 AND CRH-14 BI-DIRECTIONALLY REGULATE MALE-  
 RT TAL COLLAGEN AND SPERM SPECIFIC PROTEIN GENES OF CAENORHABDITIS  
 RT ELEGANS.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U53155; AAC48270.1; -;  
 DR EMBL: AB072926; BAB69889.1; -;  
 DR WormPep; ZC513.8; CE07615.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR002486; Col\_cuticle\_N.  
 DR InterPro; IPR002965; P\_rich\_extensin.  
 DR Pfam; PF01391; Collagen; 2\_  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 DR PRINTS; PR01217; PRICHEXTENSIN.  
 KM Hypothetical protein.  
 SQ SEQUENCE 303 AA; 29454 MW; 4D036266F31097EB CRC64;

Query Match 2.7%; Score 7; DB 5; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 GQAGTGP 262  
 Db 274 GQAGTGP 280

## RESULT 54

ID 017035 PRELIMINARY; PRT; 314 AA.  
 AC 017035;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical 30.5 kDa protein.  
 GN T15B7.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;

RA "SEQUENCE FROM N.A."  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode *C. elegans*: a platform for  
 RT investigating biology. The *C. elegans* Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]

RA "SEQUENCE FROM N.A."  
 RC STRAIN=Bristol N2;  
 RA Pauley A., Gattung S.;  
 RT "The sequence of *C. elegans* cosmid T15B7."  
 RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RA "SEQUENCE FROM N.A."  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF022985; AAB69960.1; -;  
 DR WormPep; T15B7.5; CE13656.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR002486; Col\_cuticle\_N.  
 DR Pfam; PF01391; Collagen; 3\_  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 314 AA; 30517 MW; 1CB91C0884314A7E CRC64;

Query Match 2.7%; Score 7; DB 5; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 GQAGTGP 262  
 Db 251 GQAGTGP 257

## RESULT 55

ID 08VQ72 PRELIMINARY; PRT; 322 AA.  
 AC 08VQ72;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Late competence protein ComGB.  
 GN COMGB.  
 OS Bacillus licheniformis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1402;  
 RN [1]

RA "SEQUENCE FROM N.A."  
 RC STRAIN=ATCC14580;  
 RA Lapidus A., Galleron N., Andersen J.T., Jorgensen P.L., Ehrlich S.D.,

RA Sorokin A.;  
 RT "Co-linear scaffold of the Bacillus licheniformis and Bacillus  
 RT subtilis genomes and its use to compare their competence genes";  
 RL FEMS Microbiol. Lett. 0:0-0(2002).  
 DR EMBL; AF459917; AAL67530.1; -;  
 DR InterPro; IPR001992; Bact\_secr\_systII.  
 DR Pfam; PF00482; GSP1I F.1;  
 DR PRINTS; PR00812; BCTERIALGSPF.  
 SQ SEQUENCE 322 AA; 37096 MW; E9125C5D4C6283AE CRC64;

Query Match 2.7%; Score 7; DB 2; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 NGHLYRE 35  
 |||||  
 Db 263 NGHLYRE 269

## RESULT 56

Q9J1B0 PRELIMINARY; PRT; 323 AA.  
 AC Q9J1B0;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Virus-cell attachment protein sigma C.  
 OS Nelson bay reovirus.  
 OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
 OX NCBI\_TaxID=118027;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Duncan R.;  
 RT "Nelson bay reovirus S1 genome segment."  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AF218360; AAF45159.1; -;  
 SQ SEQUENCE 323 AA; 34059 MW; 2DE9F780AE9AC358 CRC64;

Query Match 2.7%; Score 7; DB 12; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 AOSGLAS 58  
 |||||  
 Db 118 AOSGLAS 124

## RESULT 57

Q8UIK1 PRELIMINARY; PRT; 326 AA.  
 AC Q8UIK1;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Iron-sulfur cluster binding protein.  
 GN ATU0292 OR AGR\_C\_502.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D.S.,  
 RA Chapman P., Clemennin J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayavin T., Levy R., Li M.-Y., McClelland E., Palmieri A.,  
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,

RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58.";  
 RL Science 294:2317-2323 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Quicoli B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Houttel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
 RA Flanagan C., Crowell C., Giron J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328 (2001).  
 DR EMBL; AE009001; AAL1314.1; -;  
 DR EMBL; AE007968; AAK86107.1; -;  
 DR InterPro; IPR001450; 4FE4S\_Ferredoxin.  
 DR Pfam; PF00037; Fer4; 1.  
 DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 326 AA; 36255 MW; 55E14B49B6FB4F7 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 PDEDPGR 79  
 |||||  
 Db 33 PDEDPGR 39

## RESULT 58

Q941W0 PRELIMINARY; PRT; 332 AA.  
 AC Q941W0;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE B1088C09.18 protein (P0446G04.1 protein).  
 GN B1088C09.18 OR P0446G04.1.  
 OS Oryza sativa (Rice), and  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoidae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530, 39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC  
 RT clone:B1088C09.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBSJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0446G04.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AP003734; BAB88110.1; -;  
 DR EMBL; AP003252; BAB89578.1; -;  
 DR Gramene; Q941W0; -;  
 SQ SEQUENCE 332 AA; 38050 MW; 56400EF4409908A2 CRC64;

Query Match 2.7%; Score 7; DB 10; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 RSEAAAV 146  
 |||||

Db 79 RSEAAV 85

## RESULT 59

Q8YWT5 PRELIMINARY; PRT; 332 AA.  
 AC Q8YWT5; 08YWT5; 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein A11509.  
 OS All509.  
 GN Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 NC NCB1\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneo T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AF003586; BAB7875.1; -;  
 DR InterPro; IPR005524; DUF318.  
 DR Pfam; PF03773; DUF318; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 332 AA; 36033 MW; B4C6F3581AE16203 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 IATGAGI 188  
 Db 221 IATGAGI 227

## RESULT 60

Q8YTX8 PRELIMINARY; PRT; 346 AA.  
 AC Q8YTX8; 08YTX8; 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Iron(III) dicitrate transport system permease protein.  
 GN ALL285.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 NC NCB1\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneo T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AF003590; BAB74284.1; -;  
 DR InterPro; IPR000522; RecCD.  
 DR Pfam; PF01032; RecCD; 1.  
 DR ProDom; PD001357; RecCD; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 346 AA; 36510 MW; ED39A2EDC42D0F3D CRC64;

Query Match 2.7%; Score 7; DB 16; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 TLPLSAF 223  
 Db 133 TLPLSAF 139

## RESULT 61

Q8XY10 PRELIMINARY; PRT; 351 AA.  
 AC Q8XY10; 08XY10; 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE RepB.  
 GN Rhizobium etli.  
 OC Rhizobium p42b.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium.  
 NC NCB1\_TaxID=29449;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=CE3;  
 RA Cevallos M.A., Izquierdo J., Porta H., Tun C., Davila G., Brom S.;  
 RT "Rhizobium etli CB3 contains at least three plasmids of the RepABC  
 family: A structural and an evolutionary analysis";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF313446; AAM89942.1; -;  
 DR InterPro; IPR003115; ParBc.  
 DR Pfam; PF02195; ParBc; 1.  
 DR SMART; SM00470; ParB; 1.  
 DR TIGRPFAM; TIGR00180; parB\_part; 1.  
 KW plasmid.  
 SQ SEQUENCE 351 AA; 38860 MW; 041CB303304830EF CRC64;

Query Match 2.7%; Score 7; DB 2; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 AAAYOPV 149  
 Db 284 AAAYOPV 290

## RESULT 62

Q9GUT1 PRELIMINARY; PRT; 355 AA.  
 AC Q9GUT1; 09GUT1; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Y5D5A.3 protein.  
 GN Y5D5A.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC084196; AAK39622.1; -;

DR WormPep; Y55D5A.3; CE27498.  
 DR InterPro; IPR003199; Chlgly\_hydrolase.  
 DR Pfam; PF02275; CBAR; 1.  
 SQ SEQUENCE 355 AA; 40567 MW; FABS1D5F49393A73 CRC64;

Query Match 2.7%; Score 7; DB 5; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 FTTEIOE 118  
 |||||  
 Db 78 FTTEIOE 84

RESULT 63

Q927B1 PRELIMINARY; PRT; 356 AA.

AC Q927B1; 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Hypothetical protein CP0795.  
 GN CP0795 OR CP1076.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;  
 RX MEDLINE=9206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey B.K., Peterson J., Ueterbach T., Berry K., Bass S.,  
 RA Liner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 pneumoniae AR39." 28:1397-1406(2000).  
 RL Nucleic Acids Res. AAD18933.1;  
 DR EMBL; AE002264; AAF38848.1;  
 DR TIGR; CP1076;  
 DR InterPro; IPR001245; Tyr kinase.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 356 AA; 39572 MW; 2B7E096789072CF3 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDLGTLG 170  
 |||||  
 Db 2 KDLGTLG 8

RESULT 64

Q9JSA2 PRELIMINARY; PRT; 356 AA.

AC Q9JSA2; 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Hypothetical protein CP0795.  
 GN CP0795.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiae; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishi K., Hatori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL; AP002547; BAA99003.1;  
 DR InterPro; IPR001245; Tyr kinase.

DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.

KW Hypothetical protein.  
 SQ SEQUENCE 356 AA; 39614 MW; BCE076377C087356 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDLGTLG 170  
 |||||  
 Db 2 KDLGTLG 8

RESULT 65

Q927A9 PRELIMINARY; PRT; 365 AA.

AC Q927A9; 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Hypothetical protein CP0797.  
 GN CP0797.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiae; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=83558;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;  
 RX MEDLINE=9206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."  
 RL Nat. Genet. 21:385-389(1999).  
 DR EMBL; AE001661; AAD18935.1;  
 DR PHC1-2DPAGE; Q927A9;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 365 AA; 38425 MW; F34BE9BCD0983BE9 CRC64;

Query Match 2.7%; Score 7; DB 16; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDLGTLG 170  
 |||||  
 Db 142 KDLGTLG 148

RESULT 66

Q9JSA2 PRELIMINARY; PRT; 365 AA.

AC Q9JSA2; 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)  
 DE Hypothetical protein CP0797.  
 GN CP0797 OR CP1074.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]

Query Match	2.7%;	Score 7;	DB 16;	Length 365;
Best Local Similarity	100.0%;	Pred. No. 1.7e+02;		
Matches	7;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

## RESULT 67

$\frac{1}{2}$

## T "Comparison of the genomes of

Nature 417:459-463 (2002)

PFam: PF00724: oxidored FMN: 1.

Query Match	2.7%;	Score 7;	DB 16;	Length 365;
Best Local Similarity	100.0%;	Pred. No. 1.7e+02;		
Matches 7;	Conservative 0;	Mismatches 0;	Indels	

Db 56 EAAAVQP 62

## RESULT 68

ID Q8D7W9 PRELIMINARY; PRT; 370 AA.

DT 01-MAR-2003 (Tremblay, 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

NG VV2002/1  
OS Vibrio

Vibrionaceae; *Vibrio*.

RN [1] -

RC STRAIN=CMCP6;

RA Choy H.E.;

DB EMBL: AE016808: AA007006-1: -  
 KL Submitted (DEC-2002) to EMBL/GenBank/DBJ databases.

SEQUENCE 370 AA; 40414 MW; 892505680C7AC832 CRC64;

matches	;	conservative	;	mismatches	;	indels	;	gaps	;
---------	---	--------------	---	------------	---	--------	---	------	---

290 ORTIZ, P. 296

## RESULT 69

ID Q9HJF4 PRELIMINARY; PRT; 378 AA.

DT 01-MAR-2001 (TREMBLE). 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

Thermophilasma acidophilum  
JAL015.  
GN

OC Alciidae; Euryarchaeota; Thermoplasmatia  
OC Thermoplasmatales; Thermoplasma

61 RN 1004-1005 [1]

RC STRAIN=DSM 1728;

RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,

"The genome sequence of the thermoacidophilic scavenger Thermoplasma

```
KL Nacdlfe 40/:508-513(2000).
DB EMBL: AT445066: CAC12144.1: -
```

```

SO SEQUENCE 378 AA; 43103 MW; D425F17CEC6DE9C2 CRC64;

```

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

```

Matches      7;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy      17  AAYGSG 23
      129  AAYGSG 135

Db

RESULT 70
Q8H482
AC Q8H482; PRELIMINARY; PRT; 379 AA.
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Putative zinc-finger protein.
GN P0450A04.26.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzaeae; Oryza.
CX NCBI_TaxID=399477;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
RT clone:P0450A04."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004274; BAC20080.1; -.
KW Zinc.
SQ SEQUENCE 379 AA; 40201 MW; 9DF80058155C374 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 379;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      139  ARSEAA 145
      206  ARSEAA 212

Db

RESULT 71
Q9BRB6
AC Q9BRB6; PRELIMINARY; PRT; 393 AA.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Similar to receptor tyrosine kinase-like orphan receptor 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; BC006374; AAH06374.1; -.
DR HSSP; P00747; ICA.
DR InterPro; IPR000024; Fz_domain.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR000001; Kringle.
DR Pfam; PF01392; Fz; 1.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00051; Kringle; 1.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00130; KR; 1.

```

```

DR PROSITE; PS50038; Fz; 1.
DR PROSITE; PS50835; IG LIKE; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
KW Glycoprotein; Immunoglobulin domain; Kinase; Kringle; Receptor.
SQ SEQUENCE 393 AA; 43825 MW; 1F93DCBBBF53855 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 393;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      67  HSYCRNP 73
      359  HSYCRNP 365

Db

RESULT 72
Q8PKE1
AC Q8PKE1; PRELIMINARY; PRT; 394 AA.
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE MFS transporter.
GN CYNX OR XAC2234.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
CX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Fuxian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camavari F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Melandis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira C.Y., Oliveira V.R.,
RA Pereira H.A., Rossi A.J., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,
RA Trindade dos Santos M., Tufti D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AB011861; AAM37087.1; -.
KW Complete proteome.
SQ SEQUENCE 394 AA; 40055 MW; 213979057F48A7A CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 394;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      185  GAGIILG 191
      312  GAGIILG 318

Db

RESULT 73
Q9EVD2
AC Q9EVD2; PRELIMINARY; PRT; 396 AA.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Glucosyl transferase.
GN LGTG.

```

```

OS Neisseria subflava.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=28449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21142520; PubMed=11208792;
RA Arking D., Tong Y., Stein D.C.;
RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae.";
RL J. Bacteriol. 183:934-941(2001).
DR EMBL; AF241526; AAC09768.1; -.
DR InterPro; IPR001296; Glyco_trans_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW transferase.
SQ SEQUENCE 396 AA; 44368 MW; 3D82429E5258A568 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 QRTTLP 220
DB 22 QRTTLP 28

```

```

RESULT 74
Q9XC57 PRELIMINARY; PRT; 398 AA.
AC Q9XC57;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Aromatic-amino-acid aminotransferase.
GN TYRB.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA103;
RX MEDLINE=99328973; PubMed=10400585;
RA Dean C.R., Franklund C.V., Retief J.D., Coyne M.J. Jr., Hatano K.,
RA Evans D.J., Pier G.B., Goldberg J.B.;
RT "Characterization of the serogroup O11 O-antigen locus of Pseudomonas
RT aeruginosa PA103.";
RL J. Bacteriol. 181:4275-4284(1999).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
DR EMBL; AF147795; AAD45270.1; -.
DR HSSP; P04693; 3TAT.
DR InterPro; IPR004839; AminoTransferf1/2.
DR InterPro; IPR000796; Aspartate_sub.
DR InterPro; IPR004838; Nitratesf_1.
DR Pfam; PF00155; aminotran_1.2; 1.
DR PRINTS; PR00799; TRANSAMINASE.
DR PROSITE; PS00105; AA_TRANSFER_CLASS_1; 1.
KW Aminotransferase; Pyridoxal phosphate; transferase.
SQ SEQUENCE 398 AA; 43375 MW; 75D2AF8P22FFA80F CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 135 NALPARS 141
DB 169 NALPARS 175

```

```

RESULT 75
Q8CMV9 PRELIMINARY; PRT; 414 AA.
ID Q8CMV9

```

```

AC Q8CMV9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Succinyl-diaminopimelate desuccinylase.
GN SE2220.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016751; AA005862.1; -.
KW Complete proteome.
SQ SEQUENCE 414 AA; 46074 MW; FE15FB5C56B8BDAB CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 QPVIGIS 153
DB 64 QPVIGIS 70

```

```

Search completed: November 25, 2003, 13:41:09
Job time : 45 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:33:54 ; Search time 41 Seconds  
(without alignments)

1018.173 Million cell updates/sec

Title: US-10-057-951-2

Perfect score: 263  
Sequence: 1 MLAMVQAFVSNMLAEAY.....PVDPQSGSTPLMGAGTPGA 263

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : A\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	263	100.0	263	20 AA05219	Kiringlei protein s
2	263	100.0	263	20 AAW8769	Human tissue-plasm
3	263	100.0	263	22 AAB00300	Human tissue-plasm
4	263	100.0	263	23 ABR40414	Human secreted pro
5	250	95.1	263	23 AAB6149	Human ORFX ORF3001
6	172	65.4	263	21 AAB3237	Human ORFX ORF3001
7	131	49.8	146	23 ABR40487	Human secreted pro
8	131	49.8	146	23 ABR40561	Human secreted pro
9	101	38.4	263	22 AAM93748	Human polypeptide,

10	66	25.1	66	22 ABG52752	Human liver peptid
11	66	25.1	66	22 ABR31905	Peptide #5411 enco
12	66	25.1	66	22 ABR23159	Protein #5158 enco
13	66	25.1	66	22 AAM58537	Human brain expres
14	66	25.1	66	22 AAM71037	Human bone marrow
15	66	25.1	66	22 AAM18800	Peptide #5234 enco
16	66	25.1	66	22 AAM31314	Peptide #5351 enco
17	66	25.1	66	22 ABG40828	Human peptid enco
18	55	20.9	55	20 AAY12397	Human 5' EST secre
19	55	20.9	55	20 AAY05220	Kiringlei protein s
20	53	20.2	56	20 AAY12615	Human 5' EST secre
21	20	7.6	39	19 AAM72641	Nervous glia cell
22	20	7.6	39	19 AAM72640	Nervous glia cell
23	10	3.8	527	19 AAM54154	t-PA mutant (N142S
24	10	3.8	527	19 AAM54157	t-PA mutant (N142S
25	8	3.0	701	15 AAR66600	Mouse L5/3 tumour
26	8	3.0	701	15 AAM14271	Mouse growth facto
27	8	3.0	716	15 AAR66601	Mouse L5/3 tumour
28	8	3.0	716	18 AAM14272	Mouse growth facto
29	8	3.0	716	20 AAY31156	Mouse macrophage
30	8	3.0	716	20 AAM2790	Mouse MSP protein.
31	7	2.7	10	24 AAE32528	West nile virus (W
32	7	2.7	43	22 ABG57210	Human liver peptid
33	7	2.7	43	22 ABR41772	Peptide #9278 enco
34	7	2.7	43	22 AAM62644	Human brain expres
35	7	2.7	43	22 AAM75460	Human bone marrow
36	7	2.7	43	22 AAM35571	Peptide #9608 enco
37	7	2.7	43	22 ABG45027	Human peptid enco
38	7	2.7	83	23 ABR99077	AP-2gamma transcri
39	7	2.7	121	22 AAU46725	Propionibacterium
40	7	2.7	133	23 ABR07491	Human ORFX protein
41	7	2.7	140	22 ABG04700	Human ORFX protein
42	7	2.7	154	22 AAG90034	Novel human diagno
43	7	2.7	154	22 AAB70881	C. glutamicum prote
44	7	2.7	181	22 AAU32517	Novel human secret
45	7	2.7	182	20 AAY35428	Chlamydia pneumoni
46	7	2.7	186	23 ABR09730	Human ORFX protein
47	7	2.7	228	20 AAY35431	Chlamydia pneumoni
48	7	2.7	237	22 AAU87373	Novel central nerv
49	7	2.7	237	22 AAU87649	Novel central nerv
50	7	2.7	237	22 AAM84513	Human immune/haema
51	7	2.7	237	22 AAU17162	Novel signal trans
52	7	2.7	258	22 AAU34184	Staphylococcus aur
53	7	2.7	272	22 AAM41810	Human polypeptide
54	7	2.7	272	22 AAM41811	Human polypeptide
55	7	2.7	282	22 ABB59873	Drosophila melano
56	7	2.7	296	22 ABG17587	Novel human diagno
57	7	2.7	300	22 AAU36580	Staphylococcus aur
58	7	2.7	302	22 ABB68531	Drosophila melano
59	7	2.7	323	20 AAU06112	Nelson Bay virus s
60	7	2.7	332	21 AAM40024	Human ORFX ORF375
61	7	2.7	332	22 AAM40024	Human polypeptide
62	7	2.7	335	22 AAU35472	Haemophilus influe
63	7	2.7	356	23 ABR90557	Chlamydia pneumoni
64	7	2.7	357	21 AAB01333	FtsZ1 polypeptide
65	7	2.7	363	22 AAU56062	Propionibacterium
66	7	2.7	365	23 ABB80566	Chlamydia pneumoni
67	7	2.7	396	22 AAU87261	Novel central nerv
68	7	2.7	399	22 AAU17285	Novel signal trans
69	7	2.7	413	22 ABG03405	Novel human diagno
70	7	2.7	413	22 AABG16425	Novel human diagno
71	7	2.7	465	22 AAU38368	Salmonella typhi c
72	7	2.7	483	24 ABR72633	Chelonius sp. chiti
73	7	2.7	484	24 ABR40757	Zea mays oil trait
74	7	2.7	485	24 ABR40829	Zea mays oil trait
75	7	2.7	485	24 ABR40834	Zea mays oil trait
76	7	2.7	489	24 ABR40756	Zea mays oil trait
77	7	2.7	497	15 AAR36740	Human tyrosine hyd
78	7	2.7	497	15 AAB50184	Human tyrosine hyd
79	7	2.7	540	23 ABB54556	Lactococcus lactis
80	7	2.7	562	22 AAG91563	C. glutamicum prote
81	7	2.7	562	22 AAB79338	Corynebacterium gl
82	7	2.7	568	16 AAR67379	E. mitrabilis ureas

83 7 2.7 575 15 AAE54220 L. lactis alpha-ace  
 84 7 2.7 620 23 ABP30044 Streptococcus poly  
 85 7 2.7 625 22 AAB82179 CryoC #2. Bacillus  
 86 7 2.7 627 22 ABG17538 Novel human diagno  
 87 7 2.7 661 22 AAM40141 Human polypeptide  
 88 7 2.7 673 24 ABPE8303 Bacillus thuringie  
 89 7 2.7 673 24 ABPE8304 B. thuringiensis C  
 90 7 2.7 673 24 ABPE8305 B. thuringiensis C  
 91 7 2.7 673 24 ABPE8306 B. thuringiensis C  
 92 7 2.7 684 23 ABG73040 Cryptococcus nodea  
 93 7 2.7 684 23 ABG73040 Cryptococcus nodea  
 94 7 2.7 700 22 AAB79448 Corynebacterium gl  
 95 7 2.7 701 22 AAU37818 Streptococcus pneu  
 96 7 2.7 706 22 AAG92265 C. glutamicum prote  
 97 7 2.7 706 22 AAB79447 Corynebacterium g1  
 98 7 2.7 768 22 ABG19413 Novel human diagno  
 99 7 2.7 800 13 AAR28608 Potential tyrosine  
 100 7 2.7 800 20 AAM83165 Rat orphan tyrosine

## ALIGNMENTS

## RESULT 1

AAV05219 standard; Protein: 263 AA.

AAV05219;

17-JUN-1999 (first entry)

Kringel1 protein sequence.

KW Kringel1; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
 KW CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 KW neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
 KW cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
 KW myocardial infarction; hypotension; hypertension; allergy; infection;  
 KW myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
 KW male pattern baldness.

OS Homo sapiens.

PN WO911788-A1.

PD 11-MAR-1999.

PF 02-SEP-1998; 98WO-US18270.

XX 01-SEP-1998; 98US-0144889.

PR 02-SEP-1997; 97US-0056032.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PI Albopone EF, Kikily KK;

DR WPI: 1999-214707/18.

DR N-PSDB; AAX28354.

XX New kringel1 polypeptides and polynucleotides

XX Claim 1; Page 31-32; 42pp; English.

XX This sequence is a Kringel1 polypeptide of the invention.

XX The kringel1 polypeptides (II) are used to screen for agonists and

XX antagonists. Agonists are used to treat subjects in need of enhanced

XX activity or expression of (II). Antagonists are used to treat subjects

XX having need to inhibit the activity or expression of (II). The methods can

XX be used to treat conditions such as cancer, inflammation, autoimmunity,

XX allergy, asthma, rheumatoid arthritis, CNS inflammation, cerebellar

XX degeneration, Alzheimer's disease, Parkinson's disease, multiple

CC neurological abnormalities, ischaemia reperfusion injury, cardiovascular  
 CC disease, kidney disease, liver disease, ischaemic injury, myocardial  
 CC infection, hypotension, hypertension, AIDS, myelodysplastic syndromes  
 CC and other haematologic abnormalities, aplastic anaemia, male pattern  
 CC baldness, and bacterial, fungal, protozoan and viral infections. The  
 CC kringel1 polypeptides may also be used to generate antibodies.  
 CC Determining the presence or absence of mutations in, and analysing for  
 CC the presence or absence of expression of, kringel1 polynucleotides can be  
 CC used to diagnose a disease or susceptibility to a disease related to  
 CC expression or activity of kringel1 proteins. The polynucleotides may also  
 CC be used for chromosome identification, and mapping.

XX Sequence 263 AA;

Query Match 100.0%; Score 263; DB 20; Length 263;

Best Local Similarity 100.0%; Pred. No. 3..5e-227;

Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAAWQAFVLSMMLAEAYGSGGCFWMDGHLHYREQTSPAPRLCLNMLDAQSGLASAP 60  
 DB 1 MLAAWQAFVLSMMLAEAYGSGGCFWMDGHLHYREQTSPAPRLCLNMLDAQSGLASAP 60  
 QY VSGAGNHSYCRNPDEDPGRCWCVSGAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
 DB VSGAGNHSYCRNPDEDPGRCWCVSGAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
 QY 61 VSGAGNHSYCRNPDEDPGRCWCVSGAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
 DB 61 VSGAGNHSYCRNPDEDPGRCWCVSGAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
 QY 121 EGFQADEVQVAPANALPARSEAAVQPIGISQVRVMSKEKKDGLGYVIGITMVT 180  
 DB 121 EGFQADEVQVAPANALPARSEAAVQPIGISQVRVMSKEKKDGLGYVIGITMVT 180  
 QY 181 IIAIGAGIILGYSYKKGKDLKQHDQKVEREMQRTTLPLSAFTNPTCEIVDEKTVVHT 240  
 DB 181 IIAIGAGIILGYSYKKGKDLKQHDQKVEREMQRTTLPLSAFTNPTCEIVDEKTVVHT 240  
 QY 241 SCPTVPDPOEGSTPLMGAGTPGA 263  
 DB 241 SCPTVPDPOEGSTPLMGAGTPGA 263

## RESULT 2

AAW87769 standard; Protein: 263 AA.

AAW87769;

29-MAR-1999 (first entry)

DE Human tissue plasminogen activator-like protease t-PALP.

XX Tissue plasminogen activator-like protease; t-PALP; human;

XX circulatory system-related disorder; blood clotting; stroke;

XX thrombosis; peripheral arterial occlusion; pulmonary embolism;

XX myocardiothrombosis; diagnosis; therapy.

OS Homo sapiens.

XX Key

XX Peptide

XX Protein

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

CC neurological abnormalities, ischaemia reperfusion injury, cardiovascular  
 CC disease, kidney disease, liver disease, ischaemic injury, myocardial  
 CC infection, hypotension, hypertension, AIDS, myelodysplastic syndromes  
 CC and other haematologic abnormalities, aplastic anaemia, male pattern  
 CC baldness, and bacterial, fungal, protozoan and viral infections. The  
 CC kringel1 polypeptides may also be used to generate antibodies.  
 CC Determining the presence or absence of mutations in, and analysing for  
 CC the presence or absence of expression of, kringel1 polynucleotides can be  
 CC used to diagnose a disease or susceptibility to a disease related to  
 CC expression or activity of kringel1 proteins. The polynucleotides may also  
 CC be used for chromosome identification, and mapping.

XX Sequence 263 AA;

Query Match 100.0%; Score 263; DB 20; Length 263;

Best Local Similarity 100.0%; Pred. No. 3..5e-227;

Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAAWQAFVLSMMLAEAYGSGGCFWMDGHLHYREQTSPAPRLCLNMLDAQSGLASAP 60  
 DB 1 MLAAWQAFVLSMMLAEAYGSGGCFWMDGHLHYREQTSPAPRLCLNMLDAQSGLASAP 60  
 QY VSGAGNHSYCRNPDEDPGRCWCVSGAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
 DB VSGAGNHSYCRNPDEDPGRCWCVSGAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
 QY 61 VSGAGNHSYCRNPDEDPGRCWCVSGAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
 DB 61 VSGAGNHSYCRNPDEDPGRCWCVSGAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
 QY 121 EGFQADEVQVAPANALPARSEAAVQPIGISQVRVMSKEKKDGLGYVIGITMVT 180  
 DB 121 EGFQADEVQVAPANALPARSEAAVQPIGISQVRVMSKEKKDGLGYVIGITMVT 180  
 QY 181 IIAIGAGIILGYSYKKGKDLKQHDQKVEREMQRTTLPLSAFTNPTCEIVDEKTVVHT 240  
 DB 181 IIAIGAGIILGYSYKKGKDLKQHDQKVEREMQRTTLPLSAFTNPTCEIVDEKTVVHT 240  
 QY 241 SCPTVPDPOEGSTPLMGAGTPGA 263  
 DB 241 SCPTVPDPOEGSTPLMGAGTPGA 263

## RESULT 2

AAW87769 standard; Protein: 263 AA.

AAW87769;

29-MAR-1999 (first entry)

DE Human tissue plasminogen activator-like protease t-PALP.

XX Tissue plasminogen activator-like protease; t-PALP; human;

XX circulatory system-related disorder; blood clotting; stroke;

XX thrombosis; peripheral arterial occlusion; pulmonary embolism;

XX myocardiothrombosis; diagnosis; therapy.

OS Homo sapiens.

XX Key

XX Peptide

XX Protein

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

```
FT Peptide 119..128 /note= "epitope-bearing region"
FT Peptide 138..147 /note= "epitope-bearing region"
FT Peptide /note= "epitope-bearing region"
FT Peptide 155..167 /note= "epitope-bearing region"
FT Peptide 193..203 /note= "epitope-bearing region"
FT Peptide 206..215 /note= "epitope-bearing region"
FT Peptide /note= "epitope-bearing region"
FT Peptide 227..237 /note= "epitope-bearing region"
FT Peptide /note= "epitope-bearing region"
FT Peptide 243..252 /note= "epitope-bearing region"
XX
XX WO9854199-A1.
XX
XX 03-DEC-1998.
XX
XX 27-MAY-1998; 98WO-US10728.
XX
XX 28-MAY-1997; 97US-0048000.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ebner R, Moore PA, Ruben SM;
XX WPI. 1999-070207/06.
XX N-PSDB; AAV99636.
XX
XX New tissue plasminogen activator-like protease - useful in the
XX diagnosis and treatment of circulatory system-related disorders
XX
XX Claim 1; Page 56-57; 76pp; English.
XX
XX This is the amino acid sequence of tissue plasminogen activator-like
XX protease (t-PALP), a novel member of the serine protease family
XX that shares sequence homology to human tissue plasminogen activator
XX (see AAW87770). The t-PALP sequence was deduced from a cDNA clone
XX (see AAV99636) derived from activated monocytes. The 2.5 kb t-PALP
XX message has also been detected in heart, brain, lung, placenta,
XX liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate,
XX testis, ovary, small intestine, colon and peripheral blood
XX leukocytes. Isolated nucleic acids encoding amino acids -21 to
XX 242, -20 to 242, 1-242, 4-63 (kringle domain) and 64-242 (protease
XX domain) of t-PALP, or encoding epitope-bearing portions of t-PALP,
XX are also claimed, as are recombinant vectors, host cells, and
XX methods for producing t-PALP polypeptides. t-PALP may be used to
XX detect and treat disorders related to the circulatory system, and
XX to identify agonists and antagonists of t-PALP activity. The
XX homology between t-PALP and tPA indicates that t-PALP may be
XX involved in the regulation of normal and abnormal clotting
XX in e.g. stroke, deep-vein thrombosis, peripheral arterial
XX occlusion, pulmonary embolism and myocardiochrombosis.
XX
XX Sequence 263 AA:
XX
XX Query Match 100.0%; Score 263; DB 20; Length 263;
XX Best Local Similarity 100.0%; Pred. No. 3.5e-227;
XX Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMORITLPLSAFTNPCEIYDEKTVVHT 240
Db 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMORITLPLSAFTNPCEIYDEKTVVHT 240
Qy 241 SQTPVDPQEGSTPLMGAGTPGA 263
Db 241 SQTPVDPQEGSTPLMGAGTPGA 263

RESULT 3
AAE00300
ID AAE00300 standard; Protein; 263 AA.
XX
XX AAE00300;
XX
XX 13-JUN-2001 (first entry)
XX
XX Human tissue-plasminogen activator-like protease (t-PALP).
XX
XX Human, tissue-plasminogen activator-like protease; t-PALP;
XX therapy; vascular disease; stroke; deep vein thrombosis; keloid; asthma;
XX arterial occlusion; blood coagulation disorder; cerebroprotective;
XX autoimmune system disorder; human immunodeficiency syndrome; cytostatic;
XX rheumatoid arthritis; graft-versus-host disease; thyroiditis; cardiac;
XX insulin dependent diabetes; inflammatory eye disease; allergy; antiviral;
XX cardiovascular disease; heart disease; arrhythmia; myocardial ischaemia;
XX hyperproliferative disorder; hypertrophic scar; neurological disease;
XX Creutzfeldt-Jakob syndrome; neurodegenerative disorder; anti-inflammatory;
XX Alzheimer's disease; Parkinson's disease; immunosuppressive; antifungal;
XX infectious disease; drug screening; gene therapy; neuroprotective;
XX cancer; ophthalmological; antibacterial; vulnerary.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Binding-site 1..165
XX /note= "Binds to FLAG polypeptide to form
XX t-PALP-FLAG fusion protein"
XX Peptide 1..21
XX /label= Signal_peptide
XX Domain 4..63
XX /label= Kringle_domain
XX Region 12..21
XX /note= "Conserved region"
XX Protein 22..263
XX /note= "Human mature tissue-plasminogen activator-like
XX protease (t-PALP); Binds to FLAG polypeptide to form
XX t-PALP-FLAG fusion protein"
XX Region 22..38
XX /note= "Conserved region"
XX Region 22..31
XX /note= "Epitope-bearing portion"
XX Region 35..44
XX /note= "Epitope-bearing portion"
XX Region 39..49
XX /note= "Epitope-bearing portion"
XX Region 50..62
XX /note= "Conserved region"
XX Region 63..84
XX /note= "Conserved region"
XX Region 64..242
XX /label= Protease_domain
XX Domain 71..81
XX /note= "Epitope-bearing portion"
XX Region 85..97
XX /note= "Conserved region"
XX Region 91..107
XX /note= "Epitope-bearing portion"
XX Region 100..118
XX /note= "Conserved region"
XX Region 119..128
XX /note= "Epitope-bearing portion"
XX Region 119..127
```



PS Claim 11; Page 755; 873bp; English.

XX The invention relates to novel human secreted proteins and the genes  
CC encoding them. Genes and proteins of the invention may be useful for  
CC preventing, treating or ameliorating medical conditions e.g. by protein  
CC or gene therapy. These conditions include cancer and hyperproliferative  
CC disorders, immune cell proliferative disorders (e.g. leukaemia),  
CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
CC infertility, placental and uterine disorders (e.g. endometriosis),  
CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
CC disease), wound healing, gastrointestinal system disorders, particularly  
CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
CC nucleic acids are also useful for chromosome identification, radiation  
CC hybrid mapping or long-range restriction mapping. The polypeptide,  
CC polynucleotide, agonist or antagonist may also be used as a food additive  
CC or preservative to increase or decrease storage capabilities, fat content  
CC or other nutritional components. The sequences given in records  
CC ABR40409-ABR40590 and ABR2464-ABR2611 represent human secreted proteins  
CC and the genes encoding them.

XX  
SQ Sequence 263 AA;

Query Match 100.0%; Score 263; DB 23; Length 263;  
Best Local Similarity 100.0%; Pred. No. 3.5e-227;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTAWQAFIVSNMLAEAYSGGCFWMDNGHLVREDQTSAPGIRCLNMLDAOSGLASAP 60  
Db 1 MLTAWQAFIVSNMLAEAYSGGCFWMDNGHLVREDQTSAPGIRCLNMLDAOSGLASAP 60

QY 61 VSGAGNHSYCRNPDEPRGPMVCYSGAGVEKRPCEDLRCPEITTSQALPAFTTEIOEAS 120  
Db 61 VSGAGNHSYCRNPDEPRGPMVCYSGAGVEKRPCEDLRCPEITTSQALPAFTTEIOEAS 120

QY 121 EGPADBEVQFAPANALPARSEAAAVQPVIGISQVRVMSKEXKDLGTLGVIGITMMVI 180  
Db 121 EGPADBEVQFAPANALPARSEAAAVQPVIGISQVRVMSKEXKDLGTLGVIGITMMVI 180

QY 181 IIAIGAGIILGYSYKRGKDLKEQHDQVCEREMQRIPLPSAFNPCEIVDEKTVVHT 240  
Db 181 IIAIGAGIILGYSYKRGKDLKEQHDQVCEREMQRIPLPSAFNPCEIVDEKTVVHT 240

QY 241 SQTPVDPQEGSTPLMGQAGTPGA 263  
Db 241 SQTPVDPQEGSTPLMGQAGTPGA 263

RESULT 5  
AAU86149  
ID AAU86149 standard; Protein: 263 AA.

AC AAU86149;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human PRO264 polypeptide.  
XX  
KM Human; PRO; benign tumour; malignant tumour; lymphoid malignancy;  
KM leukaemia; neuronal disorder; stromal disorder; blastocoeleic disorder;  
KM inflammatory disorder; immune disorder; angiogenic disorder;  
KM cytosaratic; neuroprotective.  
XX  
OS Homo sapiens.  
XX  
PN WO200153486-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 11-FEB-2000; 2000WO-US03565.

XX  
PR 08-MAR-1999; 99WO-US05028.  
PR 11-MAR-1999; 99US-123972P.  
PR 11-MAY-1999; 99US-133459P.  
PR 02-JUN-1999; 99WO-US12022.  
PR 22-JUN-1999; 99US-140650P.  
PR 22-JUN-1999; 99US-140650P.  
PR 20-JUL-1999; 99US-144788P.  
PR 26-JUL-1999; 99US-145698P.  
PR 28-JUL-1999; 99US-146222P.  
PR 17-AUG-1999; 99US-149395P.  
PR 31-AUG-1999; 99US-151689P.  
PR 01-SEP-1999; 99WO-US20111.  
PR 15-SEP-1999; 99WO-US21090.  
PR 30-NOV-1999; 99WO-US28313.  
PR 01-DEC-1999; 99WO-US28301.  
PR 01-DEC-1999; 99WO-US28634.  
PR 05-JAN-2000; 2000WO-US00219.

XX  
PA (GENTH ) GENENTECH INC.  
PI Ashkenazi A.; Goddard A.; Godowski P.; Gurney A.; Hillan KJ;  
PI Marsters SA.; Pan J.; Pitti RM.; Roy MB.; Smith V.; Stone DM;  
PI Watanabe CK.; Wood WI;  
PI WPI; 2002-205567/26.  
DR N-PSDB; ABR40275.  
XX  
PT Thirty five nucleic acids encoding PRO polypeptides, useful for  
PT treating benign or malignant tumours, leukaemias and lymphoid  
PT malignancies, inflammatory, angiogenic and immunologic disorders -  
PS Claim 61; Fig 44; 302pp; English.

XX  
CC The present invention relates to the isolation of novel human PRO  
CC polypeptides and the polynucleotide sequences encoding them. The  
CC PRO polypeptides, agonists, antagonists or anti-PRO antibodies are  
CC useful for treating benign or malignant tumours (e.g. renal, kidney,  
CC bladder, breast, etc), leukaemias and lymphoid malignancies, other  
CC disorders such as neuronal, glial, astrocytal, hypothalamic, glandular,  
CC macrophagal, stromal and blastocoeleic disorders, inflammatory, immune  
CC and angiogenic disorders. The polynucleotide sequences are also  
CC useful in gene therapy. AAU86128-AAU86162 represent the human PRO  
CC polypeptides of the invention.

XX  
SQ Sequence 263 AA;

Query Match 95.1%; Score 250; DB 23; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1.5e-215;  
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTAWQAFIVSNMLAEAYSGGCFWMDNGHLVREDQTSAPGIRCLNMLDAOSGLASAP 60  
Db 1 MLTAWQAFIVSNMLAEAYSGGCFWMDNGHLVREDQTSAPGIRCLNMLDAOSGLASAP 60

QY 61 VSGAGNHSYCRNPDEPRGPMVCYSGAGVEKRPCEDLRCPEITTSQALPAFTTEIOEAS 120  
Db 61 VSGAGNHSYCRNPDEPRGPMVCYSGAGVEKRPCEDLRCPEITTSQALPAFTTEIOEAS 120

QY 121 EGPADBEVQFAPANALPARSEAAAVQPVIGISQVRVMSKEXKDLGTLGVIGITMMVI 180  
Db 121 EGPADBEVQFAPANALPARSEAAAVQPVIGISQVRVMSKEXKDLGTLGVIGITMMVI 180

QY 181 IIAIGAGIILGYSYKRGKDLKEQHDQVCEREMQRIPLPSAFNPCEIVDEKTVVHT 240  
Db 181 IIAIGAGIILGYSYKRGKDLKEQHDQVCEREMQRIPLPSAFNPCEIVDEKTVVHT 240

QY 241 SQTPVDPQEG 250  
Db 241 SQTPVDPQEG 250

RESULT 6

AAB43237  
ID AAB43237 standard; Protein; 263 AA.  
XX AAB43237;  
AC  
XX  
XX  
DT 08-FEB-2001 (first entry)  
XX  
XX  
DE Human CRFX ORF3001 polypeptide sequence SEQ ID NO:6002.  
XX  
KW Human; open reading frame; CRFX; detection; cytosolic; hepatoprotective;  
KW vulnerey; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antineoplastic disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN MO20005473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US06821.  
XX  
PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
DR WPI; 2000-602362/57.  
DR N-PSDB; AAC77446.  
XX  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX  
PS Claim 11; Page 5181-5182; 5507pp; English.  
XX  
XX  
CC AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The CRFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnerey;  
CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antineoplastic; antibacterial; antiviral; antifungal; antineumatic;  
CC antihypoid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 263 AA;

Query Match 65.4%; Score 172; DB 21; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1e-145;  
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 79 GPMCTVSGEAGVPEKPCEDLRCPETTSQALPATTETIQEASESGADEVQFAPANALP 138  
DB 79 GPMCTVSGEAGVPEKPCEDLRCPETTSQALPATTETIQEASESGADEVQFAPANALP 138  
QY 139 ARSEAAVQPIGTSORFMSKKEKDLGTGYVIGITMVIILAIAGAGILGYSYKRGK 198  
DB 139 ARSEAAVQPIGTSORFMSKKEKDLGTGYVIGITMVIILAIAGAGILGYSYKRGK 198  
QY 199 DLKEQHDKVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEG 250  
DB 199 DLKEQHDKVCEREMORITPLSAFTNPTCEIVDEKTVVHTSQTVPDPOEG 250  
RESULT 7  
ABR40487  
ID ABR40487 standard; Protein; 146 AA.  
XX  
XX  
AC ABR40487;  
XX  
DT 13-JUN-2003 (first entry)  
XX  
XX  
DE Human secreted protein #SEQ ID 237.  
XX  
KW Human; secreted protein; anti-HIV; neurotropic; neuroprotective;  
KW antitumoral; immunosuppressive; immunomodulator; cytosolic; cardiant;  
KW hepatotropic; antineoplastic; antiallergic; antidiabetic;  
KW gastroenteric; antineoplastic; nephrotropic; virocid; hypotensive;  
KW vasotropic; dermatological; osteopathic; antiarthritic; antiparkinsonian;  
KW antianemic; antiparasitic; cerebroprotective; antibacterial;  
KW fungicide; antiparasitic; protein therapy; gene therapy; cancer;  
KW hyperproliferative disorder; leukaemia; autoimmune disorder;  
KW immunodeficiency; HIV; infertility; uterine disorders; endometriosis;  
KW amenorrhea; ocular disorder; neurological disorder; wound healing;  
KW Huntington's disease; gastrointestinal disorder; inflammatory disease;  
KW Crohn's disease; kidney disorder; renal failure; diabetes; obesity;  
KW cirrhosis; hepatitis; cardiovascular disorder; angina; osteoarthritis;  
KW rheumatoid arthritis; psoriasis; multiple sclerosis; asthma;  
KW respiratory disorder; infectious disease; chromosome identification;  
KW food additive; nutrition.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200268628-A1.  
XX  
PD 06-SEP-2002.  
XX  
PF 21-FEB-2002; 2002WO-US05301.  
XX  
XX  
PR 23-FEB-2001; 2001US-270625P.  
PR 12-JUL-2001; 2001US-304417P.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Komatsu G, Baker KP, Fiscella M, Moore PA, Wei P;  
PI Duan DR, Shi Y, Gupta R;  
XX  
DR WPI; 2002-750417/81.  
DR N-PSDB; ABZ82542.  
XX  
PT New human secreted proteins and nucleic acids, useful for preventing,  
PT treating, or ameliorating a medical condition, e.g. cancer; autoimmune  
PT disorder, HIV, Parkinson's disease, Alzheimer's disease, diabetes,  
PT obesity or cirrhosis -  
XX  
PS Claim 11; Page 800; 873pp; English.  
XX  
XX  
CC The invention relates to novel human secreted proteins and the genes  
CC encoding them. Genes and proteins of the invention may be useful for

CC preventing, treating or ameliorating medical conditions e.g. by protein  
CC or gene therapy. These conditions include cancer and hyperproliferative  
CC disorders, immune cell proliferative disorders (e.g. leukaemia),  
CC autoimmune disorders, immunodeficiencies (e.g. HIV infection),  
CC infertility, placental and uterine disorders (e.g. endometriosis),  
CC amenorrhea, ocular disorders, neurological disorders (e.g. Huntington's  
CC disease), wound healing, gastrointestinal system disorders, particularly  
CC inflammatory diseases (e.g. Crohn's disease), kidney disorders (e.g.  
CC renal failure), diabetes, obesity, cirrhosis, hepatitis, cardiovascular  
CC disorders (e.g. angina), rheumatoid arthritis, osteoarthritis, psoriasis,  
CC multiple sclerosis, respiratory disorders (e.g. asthma), or infectious  
CC diseases caused by bacterial, parasitic, viral or fungal agents. The  
CC nucleic acids are also useful for chromosome identification, radiation  
CC hybrid mapping or long-range restriction mapping. The polypeptide,  
CC polynucleotide, agonist or antagonist may also be used as a food additive  
CC or preservative to increase or decrease storage capabilities, fat content  
CC or other nutritional components. The sequences given in records  
CC ABR40409-ABR40590 and ABZ82464-ABZ82611 represent human secreted proteins  
CC and the genes encoding them.

XX Sequence 146 AA;

XX SQ

Query Match 49.8%; Score 131; DB 23; Length 146;  
Best Local Similarity 100.0%; Pred. No. 3.1e-109; Mismatches 0; Gaps 0;  
Matches 131; Conservative 0; Indels 0;

QY 1

MLIAWVQAFIVSNMLLAAYSGGCFWMDNGHLYREDQTSAPAGRLCLMWLDAQSLASAP 60

1 MLIAWVQAFIVSNMLLAAYSGGCFWMDNGHLYREDQTSAPAGRLCLMWLDAQSLASAP 60

61 VSGAGNHSYCNPNDEDPGPGWCYVSGEAGVEKRPCEDLRCPEITTSQALPAFTTIOEAS 120

61 VSGAGNHSYCNPNDEDPGPGWCYVSGEAGVEKRPCEDLRCPEITTSQALPAFTTIOEAS 120

QY 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

EGPGADEVQVF 131

121 EGPGADEVQVF 131

Db 121

```

PN  EP1130094-A2.
XX
XX  05-SEP-2001.
PD
XX
XX  07-JUL-2000; 2000EP-0114089.
PF
XX
XX  08-JUL-1999; 99JP-0194486.
PR  11-JAN-2000; 2000JP-0118774.
PR  02-MAY-2000; 2000JP-0183765.
XX
XX  (HELI-) HELIX RES INST.
PA
XX  Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI  Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX  WPI; 2001-524255/58.
DR  N-PSDB; AAK94700.
XX
XX  830 Primers useful for synthesizing full length cDNA clones and their
PT  use in genetic manipulation -
XX
XX  Claim 8; SEQ ID NO 3727; 1380bp + sequence listing; English.
PS
XX  The invention relates to primers for synthesizing full length cDNA
CC  clones. 830 cDNA molecules encoding a human protein have been
CC  isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC  molecules have been determined. Primers for synthesizing the full length
CC  cDNA are useful for clarifying the function of the protein encoded by
CC  the cDNA. The full length clones were obtained by construction of full
CC  length enriched cDNA libraries that were synthesised by the oligo-capping
CC  method. The primers enable the production of the full length cDNA easily
CC  without any special methods. The present sequence is a polypeptide
CC  encoded by a full length human cDNA of the invention.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in CD-ROM format directly from EPO.
XX
XX  Sequence 263 AA;
SQ
XX
XX  Query Match 38.4%; Score 101; DB 22; Length 263;
XX  Best Local Similarity 100.0%; Pred. No. 3.8e-82;
XX  Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY  163 KKDGLGYLVGITMWTIIAIGAGIIGSYKRGKDLKEQHDQKCEKEMQRTITPLSA 222
DB  163 KKDGLGYLVGITMWTIIAIGAGIIGSYKRGKDLKEQHDQKCEKEMQRTITPLSA 222
QY  223 FTNPTCEIVDEKTVVHTSQTPVDPQSGSTPLMGQAGTPGA 263
DB  223 FTNPTCEIVDEKTVVHTSQTPVDPQSGSTPLMGQAGTPGA 263

RESULT 10
ABG52752
ID  ABG52752 standard; Peptide; 66 AA.
XX
XX  ABG52752;
AC
XX
XX  25-FEB-2003 (first entry)
DT
XX
XX  Human liver peptide, SEQ ID No 31400.
DE
XX
XX  Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KM  hypercholesterolaemia; coronary heart disease.
XX
XX  Homo sapiens.
OS
XX
XX  WO200157273-A2.
PN
XX
XX  09-AUG-2001.
PD
XX
XX  30-JAN-2001; 2001WO-US00664.
PF
XX
XX  04-FEB-2000; 2000US-0180312.
PR

```

```

PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX  WPI; 2001-488698/53.
DR
XX
XX  Human genome-derived single exon nucleic acid probes useful for
PT  analysing gene expression in human adult liver -
XX
XX  Claim 27; SEQ ID No 31400; 658bp; English.
PS
XX  The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC  measuring human gene expression in a sample derived from human adult
CC  liver, comprising one of 13109 defined nucleotide sequences given in the
CC  specification (or complements/ fragments). The probe hybridises at high
CC  stringency to a nucleic acid molecule expressed in the human adult
CC  liver. (II) may be used for predicting, measuring and displaying gene
CC  expression in samples derived from human adult liver. The genes
CC  identified may be involved in genetic liver diseases such as cirrhosis,
CC  hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC  is associated with coronary heart disease. ABG47348-ABG59930 represent
CC  human liver single exon encoded peptides of the invention.
CC  Note: The sequence information for this patent does not appear in the
CC  printed specification but was obtained in electronic format directly
CC  from WPI at ftp.wipo.int/pub/published_pcc_sequences.
XX
XX  Sequence 66 AA;
SQ
XX
XX  Query Match 25.1%; Score 66; DB 22; Length 66;
XX  Best Local Similarity 100.0%; Pred. No. 2.4e-51;
XX  Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY  104 TTSQALPAFTTEIQEASSEGADQVFPANALPARSEAAVQPIGISQVRNMSKEK 163
DB  1 TTSQALPAFTTEIQEASSEGADQVFPANALPARSEAAVQPIGISQVRNMSKEK 60
QY  164 KDGLTL 169
DB  61 KDGLTL 66

RESULT 11
ABB37905
ID  ABB37905 standard; Peptide; 66 AA.
XX
XX  ABB37905;
AC
XX
XX  04-FEB-2002 (first entry)
DT
XX
XX  Peptide #5411 encoded by human foetal liver single exon probe.
DE
XX
XX  Human; foetal liver; gene expression; single exon nucleic acid probe.
OS
XX
XX  Homo sapiens.
PN
XX
XX  WO200157277-A2.
PD
XX
XX  09-AUG-2001.
PF
XX
XX  30-JAN-2001; 2001WO-US00669.
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.

```



	Query Match	25.1%	Score 66;	DB 22;	Length 66;
	Best Local Similarity	100.0%;	Pred. No.	2.4e-51;	
	Matches	66;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Cy	104 TTSQLPAPFTTETICASBEGCADEYQVAPAANALPARSEAAAVPVITGISQVRNNSKEX	163                               			
Dd	1 TTSQALPAFTTEIICASBEGCAGADEYQVAPAPANALPARSEAAAVCVITGISQVRNNSKEX	60                               			
Qy	164 KDLGTL	169                   			
Db	61 KDLGTL	66             			

RESULT 12	
ABB23159	
ID	ABB23159 standard; Protein; 66 AA.
XX	
AC	ABB23159;
XX	
DT	23-JAN-2002 (first entry)
XX	
DE	Protein #5158 encoded by probe for measuring heart cell gene expression.
XX	
KW	Human; gene expression; heart; microarray; vascular system;
XX	
KW	cardiovascular disease; hypertension; cardiac arrhythmia;
XX	
KW	congenital heart disease.
XX	
OS	Homo sapiens.
XX	
PN	WO200157274-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US00666.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632386.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
PI	Penn SG, Hanzel DK, Chen W, Rank DR,
XX	

```

Query Match      25.1%; Score 66; DB 22; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.4e-51;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      104 TTSQLPAFTTEIOEASGPGADVEOVAFAPANALPARSEAAAVPVIGISQVRNMSKEX 163
      |||||
Db       1 TTSQLPAFTTEIOEASGPGADVEOVAFAPANALPARSEAAAVPVIGISQVRNMSKEX 60
      |||||

QY      164 KDLGTL 169
      |||||
Db       61 KDLGTL 66

```

Accession	Gene	Protein	Sequence	Position	Length	Score	Expect	Ident	Gap	Frame	Start	Stop	Strand	Source	Species	Organism	Accession	Gene	Protein	Sequence	Position	Length	Score	Expect	Ident	Gap	Frame	Start	Stop	Strand	Source	Species	Organism
AA058537	standard	Protein; 66 AA.																															
AA058537																																	
05-NOV-2001	(first entry)																																
Human brain	expressed single exon probe	encoded protein SEQ ID NO: 30642																															
Human	brain expressed exon; gene expression analysis; probe;																																
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;																																	
epilepsy; cancer.																																	
Homo sapiens.																																	
MO200157275-A2.																																	
09-AUG-2001.																																	
30-JAN-2001; 2001MO-US00667.																																	
04-FEB-2000; 2000US-0180312.																																	
26-MAR-2000; 2000US-0207456.																																	
30-JUN-2000; 2000US-0608408.																																	
03-AUG-2000; 2000US-0632366.																																	
21-SEP-2000; 2000US-0234687.																																	
27-SEP-2000; 2000US-0236359.																																	
04-OCT-2000; 2000GB-0024263.																																	
(MOLE-) MOLECULAR DYNAMICS INC.																																	
Penn SG, Hanzel DK, Chen W, Rank DR;																																	
WPI; 2001-483446/52.																																	

XX Example 4; SEQ ID NO: 30642; 650pp + Sequence Listing; English.  
PS  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
SQ Sequence 66 AA;  
Query Match 25.1%; Score 66; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.4e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 104 TTGQALPFTTIEQASRGPGADDEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 163  
DB 1 TTGQALPFTTIEQASRGPGADDEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
QY 164 KDLGTL 169  
DB 61 KDLGTL 66  
RESULT 14  
AAM71037  
ID AAM71037 standard; Protein; 66 AA.  
XX  
AC AAM71037;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31343.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488900/53.  
DR Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
XX Example 4; SEQ ID NO: 31343; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.  
XX

SQ Sequence 66 AA;  
Query Match 25.1%; Score 66; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.4e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 104 TTGQALPFTTIEQASRGPGADDEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 163  
DB 1 TTGQALPFTTIEQASRGPGADDEVQVFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60  
QY 164 KDLGTL 169  
DB 61 KDLGTL 66  
RESULT 15  
AAM18800  
ID AAM18800 standard; Protein; 66 AA.  
XX  
AC AAM18800;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #5234 encoded by probe for measuring cervical gene expression.  
XX  
KW Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00670.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488901/53.  
DR Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
XX Claim 27; SEQ ID No 23626; 487pp; English.  
XX  
CC The present invention relates to human single exon nucleic acid probes  
CC (SENPs: see A4110068-A418459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 66 AA;  
Query Match 25.1%; Score 66; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.4e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      104  TTGQALPAFTTEIOEASBEGADVEQVAFAPANALPARSEAAAVQPVIGISQVRMNSKEK 163
DB      1  TTGQALPAFTTEIOEASBEGADVEQVAFAPANALPARSEAAAVQPVIGISQVRMNSKEK 60
QY      164  KDLGTL 169
DB      61  KDLGTL 66

RESULT 16
AAM31314
ID  AAM31314 standard; Protein; 66 AA.
XX
AC  AAM31314;
XX
DT  17-OCT-2001 (first entry)
XX
DE  Peptide #5351 encoded by probe for measuring placental gene expression.
XX
KW  Probe; microarray; human; placenta; antenatal diagnosis;
KW  genetic disorder.
XX
OS  Homo sapiens.
XX
EN  WO200157272-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00663.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2001-488897/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
XX  analyzing gene expression in human placenta -
XX
PS  Claim 27; SEQ ID No 31583; 654bp; English.
XX
CC  The present invention relates to single exon nucleic acid probes (SENP:
CC  see A1131315-A157546). The present sequence is a peptide encoded by one
CC  such probe. The probes are useful for producing a microarray for
CC  predicting, measuring and displaying gene expression in samples derived
CC  from human placenta. The probes are useful for antenatal diagnosis of
CC  human genetic disorders.
XX
SQ  Sequence 66 AA;

Query Match 25.1%; Score 66; DB 22; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.4e-51;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

ID  ABG40828 standard; Peptide; 66 AA.
XX
AC  ABG40828;
XX
DT  19-AUG-2002 (first entry)
XX
DE  Human peptide encoded by genome-derived single exon probe SEQ ID 30493.
XX
KW  Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW  chronic obstructive pulmonary disease; interstitial lung disease;
KW  familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW  tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW  Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
KW  pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW  pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW  primary ciliary dyskinesia; pulmonary hypertension;
KW  hyaline membrane disease.
XX
OS  Homo sapiens.
XX
EN  WO200186003-A2.
XX
PD  15-NOV-2001.
XX
PF  30-JAN-2001; 2001WO-US00665.
XX
PR  04-FEB-2000; 2000US-180312P.
PR  26-MAY-2000; 2000US-207456P.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-234687P.
PR  27-SEP-2000; 2000US-236359P.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2002-114183/15.
XX
PT  Spatially-addressable set of single exon nucleic acid probes, used to
XX  measure gene expression in human lung samples -
XX
PS  Claim 27; SEQ ID No 30493; 634bp; English.
XX
CC  The invention relates to a spatially-addressable set of single exon
CC  nucleic acid probes for measuring gene expression in a sample derived
CC  from human lung comprising single exon nucleic acid probes having one of
CC  12614 nucleic acid sequences mentioned in the specification, or their
CC  complements or the 12387 open reading frames derived from the 12614
CC  probes. Also included are a microarray comprising the novel set of
CC  probes; the novel set of probes which hybridise at high stringency to a
CC  nucleic acid expressed in the human lung; measuring gene expression in a
CC  sample derived from human lung, comprising (a) contacting the array with
CC  a collection of detectably labeled nucleic acids derived from human lung
CC  mRNA, and (b) measuring the label detectably bound to each probe of
CC  the array; identifying exons in a eukaryotic genome, comprising
CC  (a) algorithmically predicting at least one exon from genomic sequences
CC  of the eukaryote; and (b) detecting specific hybridisation of
CC  labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC  comprising (a) identifying exons from genomic sequence by the method
CC  above and (b) measuring the expression of each of the exons in several
CC  tissues and/or cell types using hybridisation to a single exon
CC  microarrays having a probe with the exon, where a common pattern of
CC  expression of the exons in the tissues and/or cell types indicates that
CC  the exons should be assigned to a single gene; a peptide comprising one
CC  of 12011 sequences, mentioned in the specification, or encoded by the
CC  probes/open reading frames (ORF). The probes are used for gene
CC  expression analysis, and for identifying exons in a gene, particularly
CC  using human lung derived mRNA and for the study of lung diseases
CC  such as asthma, lung cancer, chronic obstructive pulmonary disease

```

CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberculous scleriosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemangiosarcoma, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a peptide/protein  
CC encoded by a single exon probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 66 AA;

Query Match 25.1%; Score 66; DB 23; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.4e-51;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TTSGALPATTETIQEASGEGADEVQVAPANALPARSEAAVQPIGISOVRNNSKEK 163  
DB 1 TTSGALPATTETIQEASGEGADEVQVAPANALPARSEAAVQPIGISOVRNNSKEK 60

QY 164 KDLGTL 169  
DB 61 KDLGTL 66

#### RESULT 18

AA12397  
ID AAY12397 standard; Protein; 55 AA.

XX AAY12397;  
XX 17-JUN-1999 (first entry)

XX Human 5' EST secreted protein SEQ ID NO:428.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
XX forensic; gene therapy; chromosome mapping; signal peptide;  
XX upstream regulatory sequence; cytokine activity; cell proliferation;  
XX differentiation; haematopoiesis regulation; tissue growth regulation;  
XX reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
XX thrombolytic; anti-inflammatory; tumour inhibition.

XX Homo sapiens.

XX OS

XX PN WO9906548-A2.

XX PD 11-FEB-1999.

XX PF 31-JUL-1998; 98WO-1B01222.

XX PR 01-AUG-1997; 97US-0905135.

XX (GSEST ) GENSET.

XX Ducleert A, Dumas Milne Edwards J, Lacroix B;

XX WPI; 1999-153776/13.

XX N-PSDB; AAX41230.

XX New nucleic acids encoding human secreted proteins - obtained from  
XX cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
XX kidney, lung, umbilical cord, placenta and colon tissue

XX Claim 27; Page 744; 824pp; English.

XX AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for  
XX human secreted proteins, and encode the proteins given in AAY12261 to  
XX AAY12514, respectively. The proteins given represent the signal peptide  
XX and an N-terminal fragment of a secreted protein. The nucleic acid  
XX sequences can be used for producing secreted human gene products. They

CC can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell  
CC proliferation/differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX Sequence 55 AA;

Query Match 20.9%; Score 55; DB 20; Length 55;  
Best Local Similarity 100.0%; Pred. No. 1.5e-41;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLIAVQAFIVSNMLIAEYGGGCFWNGHLYRBDQSPAPGLACIMWLDAQSG 55  
DB 1 MLIAVQAFIVSNMLIAEYGGGCFWNGHLYRBDQSPAPGLACIMWLDAQSG 55

#### RESULT 19

AA05220  
ID AAY05220 standard; Protein; 286 AA.

XX AAY05220;

XX 17-JUN-1999 (first entry)

XX Kringlet protein sequence.

XX Kringlet; cancer; inflammation; autoimmunity; rheumatoid arthritis; AIDS;  
XX CNS inflammation; cerebellar degeneration; Alzheimer's disease; asthma;  
XX Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
XX neurological abnormality; ischaemia reperfusion injury; ischaemic injury;  
XX cardiovascular disease; kidney disease; liver disease; aplastic anaemia;  
XX myocardial infarction; hypertension; hypertension; allergy; infection;  
XX myelodysplastic syndrome; haematologic abnormality; diagnosis; therapy;  
XX male pattern baldness.

XX Homo sapiens.

XX OS

XX PN WO9911788-A1.

XX PD 11-MAR-1999.

XX PF 02-SEP-1998; 98WO-US18270.

XX PR 01-SEP-1998; 98US-0144889.

XX PR 02-SEP-1997; 97US-0056032.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Albane EF, Kikly KK;

XX WPI; 1999-214707/18.

XX N-PSDB; AAX28355.

XX New kringlet polypeptides and polynucleotides

XX Claim 14; Page 33; 42pp; English.

XX This sequence is a Kringlet polypeptide of the invention.  
XX The kringlet polypeptides (I) are used to screen for agonists and  
XX antagonists. Agonists are used to treat subjects in need of enhanced  
XX activity or expression of (I). Antagonists are used to treat subjects  
XX having need to inhibit the activity or expression of (I). The methods can  
XX be used to treat conditions such as cancer, inflammation, autoimmunity,  
XX allergy, asthma, rheumatoid arthritis, CNS inflammation, cerebellar  
XX degeneration, Alzheimer's disease, Parkinson's disease, multiple

CC sclerosis, amyotrophic lateral sclerosis, head injury damage and other  
CC neurological abnormalities, ischaemia reperfusion injury, cardiovascular  
CC disease, kidney disease, liver disease, ischaemic injury, myocardial  
CC infection, hypotension, hypertension, AIDS, myelodysplastic syndromes  
CC and other haematologic abnormalities, aplastic anaemia, male pattern  
CC baldness, and bacterial, fungal, protozoan and viral infections. The  
CC kringel polypeptides may also be used to generate antibodies.  
CC Determining the presence or absence of mutations in, and analysing for  
CC the presence or absence of expression of, kringel polynucleotides can be  
CC used to diagnose a disease or susceptibility to a disease related to  
CC expression or activity of kringel proteins. The polynucleotides may also  
CC be used for chromosome identification, and mapping.

SQ Sequence 286 AA;

Query Match 20.9%; Score 55; DB 20; Length 286;  
Best Local Similarity 100.0%; Pred.No. 6.1e-41;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MLAWVQAFIVSNMLAEAYSGGCFMNDGHLHYREDQTSPPAPGLRCLNWLDAQSG 55  
DB 1 MLAWVQAFIVSNMLAEAYSGGCFMNDGHLHYREDQTSPPAPGLRCLNWLDAQSG 55

RESULT 20  
ID AAY12615  
AAV12615 standard; Protein; 56 AA.

AC AAV12615;  
DT 22-JUN-1999 (first entry)  
XX Human 5' EST secreted protein SEQ ID NO: 280 from WO 9906553.  
XX

KM Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KM forensic; gene therapy; chromosome mapping; signal peptide;  
KM upstream regulatory sequence; cytokine activity; cell proliferation;  
KM differentiation; haematopoiesis regulation; tissue growth regulation;  
KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KM thrombolytic; antiinflammatory; tumour inhibition; antitumour.

OS Homo sapiens.  
XX  
XX WO9906553-A2.  
XX  
XX 11-FEB-1999.  
XX  
XX 31-JUN-1998; 98WO-IB01237.  
XX  
XX 01-AUG-1997; 97US-0905051.  
XX  
XX (GIST ) GENSET.

XX Ductert A, Dumas Mline Edwards J, Lacroix B;  
XX WPI; 1999-153783/13.  
XX  
XX N-PSDB; AAX41473.

PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries derived from umbilical cord, lymph ganglia,  
PT lymphocytes and placental tissue

XX Claim 34; Page 376; 411pp; English.

---

CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ ligand activity, antiinflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell.

SQ Sequence 56 AA;

Query Match 20.2%; Score 53; DB 20; Length 56;  
Best Local Similarity 100.0%; Pred.No. 9.2e-40;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MLAWVQAFIVSNMLAEAYSGGCFMNDGHLHYREDQTSPPAPGLRCLNWLDAQ 53  
DB 1 MLAWVQAFIVSNMLAEAYSGGCFMNDGHLHYREDQTSPPAPGLRCLNWLDAQ 53

RESULT 21  
ID AAW72641  
AAW72641 standard; peptide; 39 AA.

AC AAW72641;  
DT 05-JAN-1999 (first entry)  
XX Nervous glia cell growth factor N-terminal peptide #2.  
XX  
XX Nervous glia cell growth factor; human; urine; secretion promoter;  
KM choline acetyltransferase activity enhancer; nervous disease.

XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 25  
FT Misc-difference /note= "unspecified"  
FT Misc-difference 29  
FT /note= "unspecified"

XX JP10265498-A.  
XX  
XX 06-OCT-1998.  
XX  
XX 24-MAR-1997; 97JP-0090305.  
XX  
XX 24-MAR-1997; 97JP-0090305.  
XX  
XX (NICH-) JAPAN CHEM RES CO LTD.  
XX

XX WPI; 1998-589719/50.

XX Nervous glia cell growth factor derived from human urine - used for  
XX treatment of nervous diseases  
XX  
XX Claim 3; Fig 7, 14pp; Japanese.

XX The present invention describes nervous glia cell growth factor, which  
XX is purified from human urine by ultrafiltration, salting-out by ammonium  
XX sulphate, gel filtration, ion exchange chromatography and reversed phase  
XX chromatography, and has a M.W. of 29 kDa by sodium dodecyl sulphate-  
XX polyacrylamide gel electrophoresis. Also described are: (1) a secretion  
XX promoter for the nerve growth factor of glia cell consisting of the  
XX above growth factor, an enhancer for choline acetyltransferase activity  
XX of neuron consisting of the above growth factor; and (2) DNA encoding  
XX acid sequence shown by the two 39 amino acid sequences as given in  
XX AAW72640 and AAW72641, which are identical, except one starts with Tyr  
XX and the other with Ser (i.e. they are from different DNA transcripts).  
XX The glia cell growth factor can be prepared in a large amount and the  
XX factor can be used for the treatment of nervous diseases.

SQ Sequence 39 AA;

Query Match 7.6%; Score 20; DB 19; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 FWDNGHL YREDQTS PAPGLR 45  
 |||||  
 DB 5 FWDNGHL YREDQTS PAPGLR 24

RESULT 22

AAW72640  
 ID AAW72640 standard; peptide; 39 AA.

XX AAW72640;

XX 05-JAN-1999 (first entry)

XX Nervous glia cell growth factor N-terminal peptide #1.

XX Nervous glia cell growth factor; human; urine; secretion promoter;

XX choline acetyltransferase activity enhancer; nervous disease.

XX Homo sapiens.

XX Key Location/Qualifiers

XX FT Misc-difference 25 /note= "unspecified"

XX FT Misc-difference 29 /note= "unspecified"

XX JPI0265498-A.

XX 06-OCT-1998.

XX 24-MAR-1997; 97JP-0090305.

XX 24-MAR-1997; 97JP-0090305.

XX (NICH-) JAPAN CHEM RES CO LTD.

XX WPI; 1998-589719/50.

XX Nervous glia cell growth factor derived from human urine - used for

XX treatment of nervous diseases

XX Claim 2; Fig 6; 14pp; Japanese.

XX The present invention describes nervous glia cell growth factor, which  
 CC is purified from human urine by ultrafiltration, salting-out by ammonium  
 CC sulphate, gel filtration, ion exchange chromatography and reversed phase  
 CC chromatography, and has a M.W. of 29 kDa by sodium dodecyl sulphate-  
 CC polyacrylamide gel electrophoresis. Also described are: (1) a secretion  
 CC promoter for the nerve growth factor of glia cell consisting of the  
 CC above growth factor, an enhancer for choline acetyltransferase activity  
 CC of neuron consisting of the above growth factor; and (2) DNA encoding  
 CC nervous glia growth factor containing a DNA sequence coding the amino  
 CC acid sequence shown by the two 39 amino acid sequences as given in  
 CC AAW72640 and AAW72641, which are identical, except one starts with Tyr  
 CC and the other with Ser (i.e. they are from different DNA transcripts).  
 CC The glia cell growth factor can be prepared in a large amount and the  
 CC factor can be used for the treatment of nervous diseases.

SQ Sequence 39 AA;

Query Match 7.6%; Score 20; DB 19; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 FWDNGHL YREDQTS PAPGLR 45  
 |||||  
 DB 5 FWDNGHL YREDQTS PAPGLR 24

RESULT 23

AAW54154  
 ID AAW54154 standard; protein; 527 AA.

XX AAW54154;

XX 20-JUL-1998 (first entry)

XX t-PA mutant (N142S).

XX Amino acid substitution; t-PA; vascular disorder; prevention;

XX fibrin deposition; adhesion formation.

XX Synthetic.

XX Key Location/Qualifiers

XX FT Domain 1..44 /note= "Finger domain"

XX FT Domain 45..91 /note= "Growth factor domain"

XX FT Domain 92..173 /note= "Kringlet-1 domain"

XX FT Domain 180..261 /note= "Kringlet-2 domain"

XX FT Domain 264..527 /note= "serine protease domain"

XX FT misc\_difference 142 /note= "N changed from wt to S in mutant"

XX US5736135-A.

XX 07-APR-1998.

XX 13-FEB-1995; 95US-0389615.

XX 11-JUL-1991; 91US-0728456.

XX 26-JAN-1993; 93US-0008940.

XX 01-APR-1994; 94US-0221660.

XX 13-FEB-1995; 95US-0389615.

XX (GETH ) GENENTECH INC.

XX Goedel DV, Leung DMH, Rice GC;

XX WPI; 1998-239153/21.

XX Mutant tissue plasminogen activator proteins - useful for treating

XX vascular disorders; preventing tissue adhesion(s), etc.

XX Claim 6; Page -: 24pp; English.

XX Mutant tissue plasminogen activator proteins (AAW54147-W54158) are  
 CC created by single or multiple amino acid substitutions. Compositions  
 CC containing the t-PA variant are used for treating vascular disorders, for  
 CC preventing fibrin deposition or for preventing adhesion formation or  
 CC reformation. Note: This sequence is not given in the specification but  
 CC was created from the wild type by the indexer.

SQ Sequence 527 AA;

Query Match 3.8%; Score 10; DB 19; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GNSHYCRNPD 74  
 |||||  
 DB 139 GNSHYCRNPD 148

RESULT 24  
 AAW54157

```

ID AA054157 standard; protein; 527 AA.
XX
AC AA054157;
XX
DT 20-JUL-1998 (first entry)
XX
DE t-PA mutant (N142S).
XX
DE t-PA mutant (N142S).
XX
KM Amino acid substitution; t-PA; vascular disorder; prevention;
KM fibrin deposition; adhesion formation.
XX
OS Synthetic.
XX
FH Key
FH Domain
FT Location/Qualifiers
FT 1..44
FT /note= "Finger domain"
FT 45..91
FT /note= "Growth factor domain"
FT 92..173
FT /note= "Kring1-1 domain"
FT 180..261
FT /note= "Kring1-2 domain"
FT 264..527
FT /note= "Serine protease domain"
FT misc_difference 142
FT /note= "N changed from wt to S in mutant"
XX
FN US5736135-A.
XX
PD 07-APR-1998.
XX
PF 13-FEB-1995; 95US-0389615.
XX
PR 11-JUL-1991; 91US-0728456.
PR 26-JAN-1993; 93US-0008940.
PR 01-APR-1994; 94US-0221660.
PR 13-FEB-1995; 95US-0389615.
XX
PA (GETH ) GENENTECH INC.
XX
PI Goeddel DV, Leung DW, Rice GC;
XX
DR WPI; 1998-239153/21.
XX
PT Mutant tissue plasminogen activator proteins - useful for treating
PT vascular disorders, preventing tissue adhesion(s), etc.
XX
PS Claim 2; Page -; 24pp; English.
XX
CC Mutant tissue plasminogen activator proteins (AA054147-W54158) are
CC created by single or multiple amino acid substitutions. Compositions
CC containing the t-PA variant are used for treating vascular disorders, for
CC preventing fibrin deposition or for preventing adhesion formation or
CC reformation. Note: This sequence is not given in the specification but
CC was created from the wild type by the indexer.
XX
SQ Sequence 527 AA;
XX
Query Match 3.8%; Score 10; DB 19; Length 527;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 65 GNHSYCRNPD 74
DB 139 GNHSYCRNPD 148

```

```

DT 13-FEB-1995 (first entry)
XX
DE Mouse L5/3 tumour suppressor protein.
XX
KM Mouse L5/3 gene; small lung cell carcinoma; tumour suppression;
KM chromosome 3; 3p21; D3F15S2 locus; hepatocyte growth factor;
KM renal cell carcinoma; Von Hippel-Lindau syndrome; predisposition;
KM kring1 domain.
XX
OS Mus musculus.
XX
FH Key
FH Peptide
FT Location/Qualifiers
FT 1..16
FT /label= signal_peptide_(16-31)
FT /note= "putative"
FT 4
FT /label= polymorphic_site
FT /note= "Pro corresponds to CCG codon in cDNA,
FT in the genomic DNA, codon 19 is CAG (Gln)"
FT 17..701
FT /note= "putative protein contains 4 kring1 domains
FT followed by a serine protease-like domain"
FT Protein
FT 57..59
FT /label= N-glycosylation_site
FT /note= "potential"
FT 158..160
FT /label= N-glycosylation_site
FT /note= "potential"
FT 290..292
FT /label= N-glycosylation_site
FT /note= "potential"
FT 605..607
FT /label= N-glycosylation_site
FT /note= "potential"
XX
FN US5315000-A.
XX
PD 24-MAY-1994.
XX
PF 14-MAY-1992; 92US-0882925.
XX
PR 14-MAY-1992; 92US-0882925.
XX
PA (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.
XX
PI Degen SJ;
XX
DR WPI; 1994-166645/20.
DR N-PSDB; AAQ79726.
XX
PT DNA from D3F15S2 locus of human chromosome 3 - encoding novel
PT growth factor, L5/3, useful as probe for detecting
PT pre-deposition towards cancer
XX
PS Disclosure; Columns 27-32; 31pp; English.
XX
XX
A mouse liver cDNA library was screened with cDNA coding for human
L5/3; the L5/3 gene, located at the D3F15S2 locus of human
chromosome 3, codes for a protein composed of 4 kring1 domains,
followed by a serine protease-like domain. The longest murine clone
to be isolated (pML5-2, AAQ79726) was not full-length. The open
reading frame was present at the 5' end of the sequence with no
CC codon for the initiator methionine in-frame with the coding
CC sequence. After determination of the sequence of the mouse gene it
CC was determined that the cDNA lacked 44bp of coding and 94bp of
CC non-coding sequence at its 5'-end.
XX
SQ Sequence 701 AA;

```

```

Query Match 3.0%; Score 8; DB 15; Length 701;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 76 DPGPWCY 83  
 |||||  
 Db 148 DPGPWCY 155

## RESULT 26

AAW14271

ID AAW14271 standard; Protein; 701 AA.

XX AAW14271;

AC 25-MAR-2003 (updated)

DT 21-JUL-1997 (first entry)

XX Mouse growth factor L5/3 partial cDNA clone ML5-2 encoded protein.

XX Mouse; growth factor; foetal; liver; probe; bovine; prothrombin; locus;  
 KW polymorphism; transition; exon; intron; chromosome; kringie domain;  
 KW cell growth; tumour suppressor; hepatocyte growth factor; regeneration.  
 XX kringie domain.  
 OS Mus musculus.

XX Key Location/Qualifiers

FH Peptide 1..16 /note= "partial signal peptide sequence"

FT Misc-difference 19 /note= "amino acid residue is Gln in this position

FT in the protein encoded by the genomic sequence  
 (AA62442); this may be due to a polymorphism"

FT Protein 17..701 /note= "mature protein"

XX US5606029-A.

XX 25-FEB-1997.

XX 18-JAN-1994; 94US-0184012.

XX 14-MAY-1992; 92US-0882925.

XX 18-JAN-1994; 94US-0184012.

XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.

XX Degen SJ;

XX WPI: 1997-163621/14.

XX N-PSDB; AA62441.

XX Human growth factor protein L5/3 - useful for altering cell growth,

XX e.g. as tumour suppressor

XX Disclosure; Column 27-32; 34pp; English.

XX This is the amino acid sequence encoded by the insert isolated from  
 CC clone ML5-2 and constitutes part of a mouse growth factor designated  
 CC L5/3 (AAW14272). The encoding sequence was isolated from a lambda-gri10  
 CC mouse liver cDNA library using a fragment of the corresponding human  
 CC cDNA sequence (AA62436) as a probe. The fragment presented here is a  
 CC partial sequence which lacks 44 bp of the coding region and 94 bp of the  
 CC non-coding region. A 1450 bp fragment of this sequence (nucleotides  
 CC 78-2188) covering 8 amino acids of the second kringie domain and all of  
 CC the third and fourth kringie domains and the serine protease-like  
 CC domain, was used to screen a mouse liver genomic DNA library to isolate  
 CC the full length genomic sequence (AA62442). The protein can be used to  
 CC alter cell growth (as a growth factor or tumour suppressor) and has  
 CC similar properties to the hepatocyte growth factor that is actively  
 CC involved in liver regeneration.  
 CC (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 701 AA;

Query Match 3.0%; Score 8; DB 18; Length 701;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 76 DPGPWCY 83  
 |||||  
 Db 148 DPGPWCY 155

## RESULT 27

AAR66601

ID AAR66601 standard; Protein; 716 AA.

XX AAR66601;

AC 14-FEB-1995 (first entry)

XX Mouse L5/3 tumour suppressor protein (from genomic sequence).

XX Mouse L5/3 gene; small lung cell carcinoma; tumour suppression;  
 KW chromosome 3; 3p21; D3F15S2 locus; hepatocyte growth factor;  
 KW renal cell carcinoma; Von Hippel-Lindau syndrome; predisposition;  
 KW kringie domain.

XX Mus musculus.

XX Key Location/Qualifiers

FH Peptide 1..31 /label= signal\_peptide

FT Misc-difference 19 /note= "putative"

FT /label= polymorphic site

FT /note= "Gln corresponds to CAG codon in genomic DNA;  
 in the cDNA, codon 19 is CCG (Pro) (AAR66600)"

FT Protein 32..716 /note= "putative protein contains 4 kringie domains  
 followed by a serine protease-like domain"

FT Modified-site 72..74 /label= N-glycosylation\_site

FT /note= "potential"

FT Modified-site 173..175 /label= N-glycosylation\_site

FT /note= "potential"

FT Modified-site 305..307 /label= N-glycosylation\_site

FT /note= "potential"

FT Modified-site 620..622 /label= N-glycosylation\_site

FT /note= "potential"

XX US5315000-A.

XX 24-MAY-1994.

XX 14-MAY-1992; 92US-0882925.

XX 14-MAY-1992; 92US-0882925.

XX (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.

XX Degen SJ;

XX WPI: 1994-166645/20.

XX N-PSDB; AAQ79727.

XX DNA from D3F15S2 locus of human chromosome 3 - encoding novel

XX growth factor, L5/3, useful as probe for detecting  
 XX pre-deposition towards cancer

XX Disclosure; Columns 33-42; 31pp; English.

XX The sequence of mouse genomic DNA coding for the L5/3 tumour  
 CC suppressor protein is composed of 18 exons separated by 17  
 CC intervening sequences. There is only one difference found between  
 CC the cDNA (AAQ79726) and genomic DNA (AAQ79727) coding sequences which



CC results in the substitution of a Gln in the gene to a Pro in the  
 CC cDNA at amino acid position 19. The putative mouse protein has the  
 CC same domain structure as its human homologue with four kringle  
 CC domains followed by a serine protease-like domain.  
 XX

Sequence 716 AA;

Query Match 3.0%; Score 8; DB 15; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 DPGPMWCY 83  
 |||||  
 Db 163 DPGPMWCY 170

RESULT 28  
 AAM14272  
 ID AAM14272 standard; Protein; 716 AA.

AC AAM14272;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 21-JUL-1997 (first entry)  
 XX

DE Mouse growth factor L5/3 complete protein.

KW Mouse; growth factor; foetal; liver; probe; bovine; prothrombin; locus;  
 KW polymorphism; transition; exon; intron; chromosome; kringle domain;  
 KW cell growth; tumour suppressor; hepatocyte growth factor; regeneration.  
 XX

OS Mus musculus.

PH Key Location/Qualifiers  
 PT Peptide 1..31  
 PT /note= "signal peptide"

PT Misc-difference 19

PT /note= "amino acid residue is Pro at this position  
 in the protein encoded by the cDNA clone MUs-2  
 (AAT62441); this may be due to a polymorphism"

PT Protein 32..716

PT /note= "mature protein"

PT Modified-site 72

PT /note= "N-linked glycosylation site"

PT Modified-site 173

PT /note= "N-linked glycosylation site"

PT Modified-site 305

PT /note= "N-linked glycosylation site"

PT Modified-site 624

PT /note= "N-linked glycosylation site"

XX US5606029-A.

XX 25-FEB-1997.

XX 18-JAN-1994; 94US-0184012.

XX 14-MAY-1992; 92US-0882925.

XX 18-JAN-1994; 94US-0184012.

XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.

XX Degen SJ;

XX WPI; 1997-153621/14.

XX N-PSDB; AAT62442.

XX Human growth factor protein L5/3 - useful for altering cell growth,  
 e.g. as tumour suppressor  
 XX Disclosure; Column 33-42; 34pp; English.  
 XX This is the amino acid sequence of the wild type mouse growth factor

CC designated L5/3. The protein sequence differs from that encoded by the  
 CC cDNA clone (AAT62441) at position 19; in this sequence a Gln, in the  
 CC cDNA clone a Pro. This difference may be due to a polymorphism at this  
 CC codon. The full length mouse gene contains 18 exons and encodes a protein  
 CC having a molecular weight 80 kD. The protein can be used to alter cell  
 CC growth (as a growth factor or tumour suppressor) and has similar  
 CC properties to the hepatocyte growth factor that is actively involved in  
 CC liver regeneration.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 XX

Sequence 716 AA;

Query Match 3.0%; Score 8; DB 18; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 DPGPMWCY 83  
 |||||  
 Db 163 DPGPMWCY 170

RESULT 29  
 AAY31156  
 ID AAY31156 standard; protein; 716 AA.

AC AAY31156;  
 XX  
 DT 26-OCT-1999 (first entry)  
 XX

DE Murine macrophage stimulating protein.

KW Macrophage stimulating protein; MSP; murine; modulator; proliferation;  
 KW differentiation; intestinal epithelium; colon crypt; treatment; cancer;  
 KW haematopoietic disorder; megakaryocyte deficiency; gastrointestinal;  
 KW chemotherapeutic agent; gut toxicity.  
 XX

OS Mus sp.

XX US5948892-A.

XX 07-SEP-1999.

XX 16-DEC-1996; 96US-0766982.

XX 16-DEC-1996; 96US-0766982.

XX (AMGE-) AMGEN INC.

XX Wahl RC;

XX WPI; 1999-517975/43.

XX Claim 4; Column 15-18; 23pp; English.

PT Analogues of macrophage stimulating protein for treating  
 PT gastrointestinal or haematopoietic disorders  
 XX

XX This invention describes a novel purified and isolated analogue of mature

XX macrophage stimulating protein (MSP) having at least one unpaired

XX cysteine residue substituted with another amino acid which modulates the

XX proliferation or differentiation of the intestinal epithelium. The

XX product of the invention binds to RON (a cell membrane protein tyrosine

XX kinase which is a member of the c-met family) to promote the formation of

XX colon crypts. MSP analogues are useful for the treatment of conditions

XX requiring the administration of MSP, such conditions include

XX haematopoietic disorders such as those involving a deficiency of

XX megakaryocytes and gastrointestinal disorders such as ulcerative colitis,

XX Crohn's disease and infections. The MSP analogues are useful for

XX maintaining and repairing the epithelial lining in the treatment of

XX cancer, where the aggressive use of chemotherapeutic agents or the use of

XX whole body radiation may lead to gut toxicity. The MSP analogues, which

XX have a higher activity than normal human MSP are effective at smaller

XX dosages, or optionally, they may be administered less frequently than

CC human MSP. This sequence represents the murine MSP described in the  
CC method of the invention.

XX  
SQ Sequence 716 AA;

Query Match 3.0%; Score 8; DB 20; Length 716;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 DPGPWCY 83  
DB 163 DPGPWCY 170

RESULT 30  
AAW82790  
ID AAW82790 standard; Protein; 716 AA.

XX AAW82790;

DT 12-APR-1999 (first entry)

DE Mouse MSP protein.

XX MSP, macrophage stimulating protein; apoptosis; murine; treatment;  
KW neuroendocrine cell; RON receptor; small cell lung carcinoma; tumour;  
KW pathogen infection; thrombocyte production; megakaryocyte maturation;  
KW thrombocytopaenia; hepatocyte growth.

OS Mus sp.

XX WO9855141-A1.

XX 10-DEC-1998.

PF 04-JUN-1998; 98WO-US11573.

PR 04-JUN-1997; 97US-0048594.

PA (BGMH) BRIGHAM & WOMENS HOSPITAL.

PI Sunday ME, Willet C;

DR WPI; 1999-059877/05.

DR N-PSDB; AAV72084.

XX Treating tumours derived from neuroendocrine cells with macrophage  
PT stimulating protein - or its nucleic acid, also for preventing  
PT development of these tumours, specifically small cell lung carcinoma  
XX  
PS Disclosure; Page 67-69; 100pp; English.

CC This sequence represents a novel murine macrophage stimulating protein,  
CC MSP, which is used in a method for the prophylactic treatment of a tumour  
CC derived from neuroendocrine cells (NEC) by administration of this MSP  
CC to a subject at risk, sufficient to induce apoptosis of NEC expressing a  
CC RON receptor (the receptor for MSP). The method is especially used to  
CC treat or prevent small cell lung carcinoma and apoptosis of  
CC RON-expressing cells may be induced in vivo or in vitro. Screening NEC  
CC from a subject for susceptibility to MSP-induced apoptosis is  
CC used to identify patients who will benefit from treatment with the MSP  
CC protein. MSP is already known for treating pathogen infections, for  
CC stimulating thrombocyte production and megakaryocyte maturation (for  
CC treating thrombocytopaenia) and for stimulating growth of cells  
CC (particularly hepatocytes).

XX Sequence 716 AA;

Query Match 3.0%; Score 8; DB 20; Length 716;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 DPGPWCY 83

DB 163 DPGPWCY 170

RESULT 31

ID AAE32528 standard; peptide; 10 AA.

XX AAE32528;

DT 24-MAR-2003 (first entry)

DE West Nile virus (WNV) peptide #24.

XX West Nile virus; WNV; flavivirus-related disorder; vaccine; virucide;  
KW immunopathogenesis; pharmacogenomic; infection; immune response.

OS West Nile virus.

PN WO200283903-A2.

PD 24-OCT-2002.

PF 28-FEB-2002; 2002WO-US06575.

PR 28-FEB-2001; 2001US-272132P.

PA (UYBR-) UNIV BROWN RES FOUND.

PI Degroot AS, Martin W;

DR WPI; 2003-103373/09.

XX New West Nile Virus (WNV) vaccine, useful for inducing an anti-WNV  
PT immune response to a mammalian subject, for treating or preventing WNV  
PT infection or flavivirus-related disorders -  
XX  
XX Claim 1; Page 8; 103pp; English.

XX The invention relates to vaccines comprising West Nile virus (WNV)  
CC vaccine candidate peptides. It also relates to methods for determining  
CC MNV epitopes and methods of using the same. The vaccine is useful for  
CC inducing an anti-WNV immune response to a mammalian subject, for  
CC treating or preventing WNV infection or flavivirus-related disorders.  
CC The candidate peptides are useful for screening exposed individuals,  
CC investigating the immunopathogenesis of WNV disease in humans, as  
CC components of diagnostic kits developed for the surveillance effort,  
CC or as tool for measuring WNV vaccine-related immune responses. The  
CC recombinant peptides and polynucleotides are useful in diagnostic  
CC assays, prognostic assays, pharmacogenomics, or for monitoring clinical  
CC trials. The present sequence is West Nile virus peptide used in the  
CC invention.

XX Sequence 10 AA;

Query Match 2.7%; Score 7; DB 24; Length 10;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 PGLRCLN 48  
DB 2 PGLRCLN 8

RESULT 32

ID ABG57210 standard; Peptide; 43 AA.

XX ABG57210;

DT 25-FEB-2003 (first entry)

DE Human liver peptide, SEQ ID No 35858.

```

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KM hypercholesterolaemia; coronary heart disease.
XX
XX Homo sapiens.
OS
XX MO200157273-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00664.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-48898/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human adult liver -
XX
XX Claim 27; SEQ ID NO 3588; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult
CC liver. (II) may be used for predicting, measuring and displaying gene
CC expression in samples derived from human adult liver. The genes
CC identified may be involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC is associated with coronary heart disease. ABG47348-ABG59330 represent
CC human liver single exon encoded peptides of the invention.
CC Note: The sequence information for this patent does not appear in the
CC printed specification but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 43 AA;
SQ
XX
XX Query Match 2.7%; Score 7; DB 22; Length 43;
XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 176 TMMVIIII 182
QY |||||
DB 26 TMMVIIII 32
XX
XX
XX RESULT 33
XX ABB41772
XX ID ABB41772 standard; Peptide; 43 AA.
XX
XX ABB41772;
AC
XX
XX 04-FEB-2002 (first entry)
DT
XX
XX Peptide #9278 encoded by human foetal liver single exon probe.
DE
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
OS
XX MO200157277-A2.
PN
XX
XX

```

```

PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483447/52.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 34407; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 43 AA;
SQ
XX
XX Query Match 2.7%; Score 7; DB 22; Length 43;
XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 176 TMMVIIII 182
QY |||||
DB 26 TMMVIIII 32
XX
XX
XX RESULT 34
XX AAM62644
XX ID AAM62644 standard; Protein; 43 AA.
XX
XX AAM62644;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 34749.
DE
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer.
XX
XX Homo sapiens.
OS
XX
XX MO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX

```

```
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 34749; 650bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
XX Sequence 43 AA;
XX
XX
XX Query Match 2.7%; Score 7; DB 22; Length 43;
XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 176 TMMVIIII 182
XX |||||
XX 26 TMMVIIII 32
XX
XX Db
XX
XX RESULT 35
XX AAM75460
XX ID AAM75460 standard; Protein; 43 AA.
XX
XX AAM75460;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 35766.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 35766; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
```

```
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
XX Sequence 43 AA;
XX
XX
XX Query Match 2.7%; Score 7; DB 22; Length 43;
XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 176 TMMVIIII 182
XX |||||
XX 26 TMMVIIII 32
XX
XX Db
XX
XX RESULT 36
XX AAM35571
XX ID AAM35571 standard; Protein; 43 AA.
XX
XX AAM35571;
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #9608 encoded by probe for measuring placental gene expression.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID NO 35840; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AI57546). The present sequence is a peptide encoded by one
XX of these probes. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX
XX Sequence 43 AA;
XX
XX
XX Query Match 2.7%; Score 7; DB 22; Length 43;
XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 176 TMMVIIII 182
XX |||||
XX 26 TMMVIIII 32
XX
XX Db
```

RESULT 37  
ABG45027  
ID ABG45027 standard; Peptide; 43 AA.  
XX  
XX ABG45027;  
AC  
AC 19-AUG-2002 (first entry)  
XX  
XX Human peptide encoded by genome-derived single exon probe SEQ ID 34692.  
DE  
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
XX Chronic obstructive pulmonary disease; interstitial lung disease;  
XX Familial idiopathic pulmonary fibrosis; neurofibromatosis;  
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;  
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
XX primary ciliary dyskinesia; pulmonary hypertension;  
XX hyaline membrane disease.  
XX  
XX Homo sapiens.  
OS  
XX WO200186003-A2.  
XX  
XX 15-NOV-2001.  
PD  
XX 30-JAN-2001; 2001WO-US00665.  
PF  
XX 04-FEB-2000; 2000US-180312P.  
XX 26-MAY-2000; 2000US-207456P.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-234687P.  
PR 27-SEP-2000; 2000US-236359P.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples -  
PT  
XX Claim 27; SEQ ID No 34692; 634bp; English.  
PS  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridise at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung; comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe;  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarray having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene

CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a peptide/protein  
CC encoded by a single exon probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX

SQ Sequence 43 AA;

Query Match 2.7%; Score 7; DB 23; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 TMWVYIII 182  
|||||  
Db 26 TMWVYIII 32

RESULT 38  
ABB99077  
ID ABB99077 standard; Protein; 83 AA.  
XX  
XX ABB99077;  
AC  
XX 07-FEB-2003 (first entry)  
DT  
XX AP-2gamma transcription factor 9.13 amino acid sequence.  
DE  
XX AP-2gamma transcription factor 9.13; embryo development disorder.  
XX  
XX AP-2gamma transcription factor 9.13; embryo development disorder.  
XX  
XX Unidentified.  
OS  
XX CN1351020-A.  
XX  
XX 29-MAY-2002.  
PD  
XX 26-OCT-2000; 2000CN-0125786.  
PF  
XX 26-OCT-2000; 2000CN-0125786.  
PR  
XX 26-OCT-2000; 2000CN-0125786.  
XX  
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
PA  
XX Mao Y, Xie Y;  
PI WPI; 2002-644401/70.  
XX  
XX N-PSDB; ABV75472.  
DR  
XX Polyptide-AP-2 gamma transcription factor 9.13 -  
XX  
XX Claim 1; Page 26-27 (disclosure); 33pp; Chinese.  
PS  
XX The invention relates to a new polypeptide designated AP-2 gamma  
CC transcription factor 9.13. The application of the polypeptide is in  
CC treating diseases such as embryo development disorder. The antagonist of  
CC the polypeptide and its medical action are also disclosed. The current  
CC sequence represents the AP-2gamma transcription factor 9.13 amino acid  
CC sequence.  
XX  
XX

SQ Sequence 83 AA;

Query Match 2.7%; Score 7; DB 23; Length 83;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDLGTLG 170  
Db 76 KDLGTLG 82

## RESULT 39

AAU46725  
ID AAU46725 standard; Protein; 121 AA.

XX  
AC AAU46725;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #7621.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeity YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR N-PSDB; AASS59535.

DR WPI; 2001-616774/71.

PT Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

XX Example 1; SEQ ID No 7920; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 121 AA;

Query Match 2.7%; Score 7; DB 22; Length 121;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 TSOALPA 111  
Db 1 TSOALPA 7

## RESULT 40

ABP07491  
ID ABP07491 standard; Protein; 133 AA.

XX  
AC ABP07491;

DT 25-JUN-2002 (first entry)

DE Human ORFX protein sequence SEQ ID NO:14964.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KM hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KM degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KM hypertension; hypothyroidism; cholesterol ester storage disease;  
KM immune deficiency; immune disorder; infectious disease;  
KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KM myasthenia gravis.

OS Homo sapiens.

PN WO200192523-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US10836.

PR 30-MAY-2000; 2000US-206132P.

PR 29-AUG-2000; 2000US-228716P.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach MD;

DR WPI; 2002-106308/14.

DR N-PSDB; ABN3243.

PT Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,

PT hyperproliferative disorders and autoimmune disorders -

XX Disclosure; SEQ ID 14964; 1037pp; English.

XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 133 AA;  
SQ Query Match 2.7%; Score 7; DB 23; Length 133;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 EGPQADE 127  
|||  
Db 101 EGPQADE 107

RESULT 41  
ABG04700  
ID ABG04700 standard; Protein; 140 AA.  
XX  
AC ABG04700;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #4691.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS68887.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 35059; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 140 AA;

XX Query Match 2.7%; Score 7; DB 22; Length 140;  
SQ Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 119 ASEGPQA 125  
|||  
Db 10 ASEGPQA 16

RESULT 42  
AAG90034  
ID AAG90034 standard; Protein; 154 AA.  
XX  
AC AAG90034;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum protein fragment SEQ ID NO: 3788.  
XX  
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
PA (KYOW) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
DR WPI; 2001-376931/40.  
DR N-PSDB; AAH65253.  
XX  
XX  
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT -  
XX  
PS Claim 17; SEQ ID NO: 3788; 246bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
SQ Sequence 154 AA;

Qy 187 GIILGYS 193  
|||  
Db 50 GIILGYS 56

```

RESULT 43
AAB70881
ID AAB70881 standard; Protein; 154 AA.
XX
AC AAB70881;
XX
DT 12-JUL-2001 (first entry)
XX
DE C. glutamicum lrp protein.
XX
KW Fermentation; L-amino acid production; lrp gene; lysine; isoleucine;
KM medicine; animal feed supplement.
XX
OS Corynebacterium glutamicum.
XX
PN EP1090993-A1.
XX
PD 11-APR-2001.
XX
PF 29-SEP-2000; 2000EP-0121159.
XX
PR 05-OCT-1999; 99DE-1047792.
XX
PA (DEGS ) DEGUSSA-HUELS AG.
XX
PI Moeckel B, Pfeifferle W, Puehler A, Kalinowski J, Bathe B;
XX
DR WPI: 2001-292927/31.
XX
DR N-PSDB; AAF61688.
XX
PT New lrp gene from coryneform bacteria, used to prepare transformants
PT with increased synthesis of amino acids, particularly lysine and
PT isoleucine -
XX
PS Claim 6; Page 15; 22pp; German.
XX
CC This invention describes a novel isolated nucleic acid (I) from
CC coryneform bacteria used for the fermentative production of selected
CC L-amino acids, by fermenting the amino acid-producing coryneform in
CC which at least the lrp gene has been weakened or amplified, then
CC isolating amino acids that have accumulated in the medium or cells. (I)
CC is used to transform coryneforms for production of L-amino acids,
CC specifically lysine and isoleucine, which are used in medicine and
CC particularly as animal feed supplement. It may also be used as probes
CC and primers for isolating related sequences. Regulating expression of
CC (I) improves production of amino acids, especially of L-lysine. This
CC sequence represents the Corynebacterium glutamicum lrp protein which is
CC used in the method described in the invention.
XX
SQ Sequence 154 AA;

Query Match 2.7%; Score 7; DB 22; Length 154;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 GILIGYS 193
DB 50 GILIGYS 56

RESULT 44
AAU32517
ID AAU32517 standard; Protein; 181 AA.
XX
AC AAU32517;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #3008.
XX
KM Human; vaccination; gene therapy; nutritional supplement;

```

```

KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
XX
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 625-626; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 181 AA;

Query Match 2.7%; Score 7; DB 22; Length 181;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 PAPGLRC 46
DB 117 PAPGLRC 123

RESULT 45
AAJ35428
ID AAJ35428 standard; Protein; 182 AA.
XX
AC AAJ35428;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae transmembrane protein sequence.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KM vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PR 03-JUN-1999.

```



```

XX 20-NOV-1998; 98WO-IB01890.
PF
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST ) GENSET.
XX
PI Grifffals R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1210; Disclosure; 1912pp; English.
XX
CC AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
CC
SQ Sequence 182 AA;
XX
Query Match 2.7%; Score 7; DB 20; Length 182;
Best local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 164 KDLGTLG 170
|||
42 KDLGTLG 48
DB
RESULT 46
ABP09730
ID ABP09730 standard; Protein; 196 AA.
XX
AC ABP09730;
XX
DT 25-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:19442.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic erythematousus;
KW hypertension; hypothyroidism; cholesterol ester scorage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PP 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach MD;
XX

```

```

DR WPI; 2002-106308/14.
DR N-PSDB; ABN25482.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders
XX
PS Disclosure; SEQ ID 19442; 1037pp; English.
XX
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP1500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantion, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 196 AA;
XX
Query Match 2.7%; Score 7; DB 23; Length 196;
Best local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 244 PVDPQEG 250
|||
27 PVDPQEG 33
DB
RESULT 47
AAY35431
ID AAY35431 standard; Protein; 228 AA.
XX
AC AAY35431;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae transmembrane protein sequence.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN MO9927105-A2.
XX
PD 03-JUN-1999.
XX
PP 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST ) GENSET.
XX

```

PI Giffais R;  
XX  
XX WPI; 1999-357842/30.  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1212; Disclosure; 1912pp; English.  
XX  
CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
SQ Sequence 228 AA;  
  
Query Match 2.7%; Score 7; DB 20; Length 228;  
Best Local Similarity 100.0%; Pred.No.4,7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 164 KDLGTG 170  
|||  
Db 158 KDLGTG 164  
  
RESULT 48  
AAU87373  
ID AAU87373 standard; Protein; 237 AA.  
AC AAU87373;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Novel central nervous system protein #283.  
XX  
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
KW myocardial infarction; wound healing; cell proliferation; skin aging;  
KW food additive; food preservative; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO20015318-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01332.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224516.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226888.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.

CC	e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders			
CC	e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,			
CC	adenocarcinomas and irritable bowel syndrome, reproductive system			
CC	disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes			
CC	and pituitary dwarfism, cancer and disorders at the cellular level e.g.			
CC	leukaemia, disorders involving neovascularisation e.g. malignancies,			
CC	respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.			
CC	acute kidney failure and blood related disorders e.g. myocardial			
CC	infarction. The polypeptides can also be used to aid wound healing and			
CC	epithelial cell proliferation, to prevent skin aging due to sunburn, to			
CC	maintain organs before transplantation, for supporting cell culture of			
CC	primary tissues, to regenerate tissues and in chemotaxis. The			
CC	polypeptides can also be used as a food additive or preservative to			
CC	increase or decrease storage capabilities, fat content, lipid, protein.			
Query Match				
Best Local Similarity 2.7%; Score 7; DB 22; Length 237;				
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	149 VIGISQR 155			
Db	215 VIGISQR 221			
RESULT 49				
AAU87649				
ID	AAU87649 standard; Protein; 237 AA.			
XX	AAU87649;			
AC	AAU87649;			
XX				
DT	05-JUN-2002 (first entry)			
XX				
DE	Novel central nervous system protein #559.			
XX				
KM	Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;			
KM	hyperproliferative disorder; neoplasm; cardiovascular disorder;			
KM	cardiac arrest; cerebrovascular disorder; ischaemia; angiosclerosis;			
KM	nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;			
KM	acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;			
KM	adenocarcinoma; reproductive system disorder; testicular feminisation;			
KM	endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;			
KM	respiratory disorder; renal disorder; kidney failure; blood disorder;			
KM	myocardial infarction; wound healing; cell proliferation; skin aging;			
XX	food additive; food preservative; gene therapy.			
OS	Homo sapiens.			
XX				
XX	WO200155318-A2.			
XX				
PD	02-AUG-2001.			
XX				
PF	17-JAN-2001; 2001WO-US01332.			
XX				
PR	31-JAN-2000; 2000US-0179065.			
PR	04-FEB-2000; 2000US-0180628.			
PR	24-FEB-2000; 2000US-0184664.			
PR	02-MAR-2000; 2000US-0186350.			
PR	16-MAR-2000; 2000US-0189874.			
PR	17-MAR-2000; 2000US-0190076.			
PR	18-APR-2000; 2000US-0198123.			
PR	19-MAY-2000; 2000US-0205515.			
PR	07-JUN-2000; 2000US-0209467.			
PR	28-JUN-2000; 2000US-0214886.			
PR	30-JUN-2000; 2000US-0215135.			
PR	07-JUL-2000; 2000US-0216647.			
PR	07-JUL-2000; 2000US-0216880.			
PR	11-JUL-2000; 2000US-0217487.			
PR	11-JUL-2000; 2000US-0217496.			
PR	14-JUL-2000; 2000US-0218290.			
PR	26-JUL-2000; 2000US-0220963.			
PR	26-JUL-2000; 2000US-0220964.			
PR	14-AUG-2000; 2000US-0224518.			
PR	14-AUG-2000; 2000US-0224519.			

PR	14-AUG-2000	2000US-0225213.	PR	08-NOV-2000	2000US-0246523.
PR	14-AUG-2000	2000US-0225214.	PR	08-NOV-2000	2000US-0246524.
PR	14-AUG-2000	2000US-0225216.	PR	08-NOV-2000	2000US-0246525.
PR	14-AUG-2000	2000US-0225267.	PR	08-NOV-2000	2000US-0246526.
PR	14-AUG-2000	2000US-0225268.	PR	08-NOV-2000	2000US-0246527.
PR	14-AUG-2000	2000US-0225270.	PR	08-NOV-2000	2000US-0246528.
PR	14-AUG-2000	2000US-0225347.	PR	08-NOV-2000	2000US-0246532.
PR	14-AUG-2000	2000US-0225357.	PR	08-NOV-2000	2000US-0246609.
PR	14-AUG-2000	2000US-0225758.	PR	08-NOV-2000	2000US-0246610.
PR	14-AUG-2000	2000US-0225759.	PR	08-NOV-2000	2000US-0246611.
PR	18-AUG-2000	2000US-0226279.	PR	08-NOV-2000	2000US-0246613.
PR	22-AUG-2000	2000US-0226681.	PR	17-NOV-2000	2000US-0249207.
PR	22-AUG-2000	2000US-0226688.	PR	17-NOV-2000	2000US-0249208.
PR	22-AUG-2000	2000US-0227182.	PR	17-NOV-2000	2000US-0249209.
PR	23-AUG-2000	2000US-0227009.	PR	17-NOV-2000	2000US-0249210.
PR	30-AUG-2000	2000US-0228924.	PR	17-NOV-2000	2000US-0249211.
PR	01-SEP-2000	2000US-0229287.	PR	17-NOV-2000	2000US-0249212.
PR	01-SEP-2000	2000US-0229343.	PR	17-NOV-2000	2000US-0249213.
PR	01-SEP-2000	2000US-0229344.	PR	17-NOV-2000	2000US-0249214.
PR	01-SEP-2000	2000US-0229345.	PR	17-NOV-2000	2000US-0249215.
PR	05-SEP-2000	2000US-0229509.	PR	17-NOV-2000	2000US-0249216.
PR	05-SEP-2000	2000US-0229513.	PR	17-NOV-2000	2000US-0249217.
PR	06-SEP-2000	2000US-0230437.	PR	17-NOV-2000	2000US-0249218.
PR	06-SEP-2000	2000US-0230438.	PR	17-NOV-2000	2000US-0249219.
PR	08-SEP-2000	2000US-0231242.	PR	17-NOV-2000	2000US-0249245.
PR	08-SEP-2000	2000US-0231243.	PR	17-NOV-2000	2000US-0249264.
PR	08-SEP-2000	2000US-0231244.	PR	17-NOV-2000	2000US-0249265.
PR	08-SEP-2000	2000US-0231413.	PR	17-NOV-2000	2000US-0249297.
PR	08-SEP-2000	2000US-0231414.	PR	17-NOV-2000	2000US-0249299.
PR	08-SEP-2000	2000US-0231415.	PR	17-NOV-2000	2000US-0249300.
PR	08-SEP-2000	2000US-0232080.	PR	01-DEC-2000	2000US-0250160.
PR	12-SEP-2000	2000US-0231968.	PR	01-DEC-2000	2000US-0250391.
PR	14-SEP-2000	2000US-0232397.	PR	05-DEC-2000	2000US-0251030.
PR	14-SEP-2000	2000US-0232398.	PR	05-DEC-2000	2000US-0251988.
PR	14-SEP-2000	2000US-0232399.	PR	05-DEC-2000	2000US-0256719.
PR	14-SEP-2000	2000US-0232400.	PR	06-DEC-2000	2000US-0251479.
PR	14-SEP-2000	2000US-0232401.	PR	08-DEC-2000	2000US-0251856.
PR	14-SEP-2000	2000US-0233063.	PR	08-DEC-2000	2000US-0251868.
PR	14-SEP-2000	2000US-0233064.	PR	08-DEC-2000	2000US-0251869.
PR	14-SEP-2000	2000US-0233065.	PR	08-DEC-2000	2000US-0251989.
PR	21-SEP-2000	2000US-0234223.	PR	08-DEC-2000	2000US-0251990.
PR	21-SEP-2000	2000US-0234274.	PR	11-DEC-2000	2000US-0254097.
PR	25-SEP-2000	2000US-0234997.	PR	05-JAN-2001	2001US-0259678.
PR	25-SEP-2000	2000US-0234998.			
PR	26-SEP-2000	2000US-0235484.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000	2000US-0235834.	XX		
PR	27-SEP-2000	2000US-0235836.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000	2000US-0236327.	XX	WPI; 2001-581633/65.	
PR	29-SEP-2000	2000US-0236367.	DR	N-PSDB; ABK43979.	
PR	29-SEP-2000	2000US-0236368.	XX		
PR	29-SEP-2000	2000US-0236369.	XX		
PR	29-SEP-2000	2000US-0236370.	PT	New isolated nucleic acid encoding a protein for diagnosing,	
PR	02-OCT-2000	2000US-0236802.	PT	preventing, treating or ameliorating medical conditions and	

CC acute kidney failure and blood related disorders e.g. myocardial  
CC infection. The polyepitopes can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 2.7%; Score 7; DB 22; Length 237;

Best Local Similarity 100.0%; Pred. No. 4.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 VIGISQR 155  
|||||

Db 215 VIGISQR 221

RESULT 50

AAM84513 standard; Protein; 237 AA.

AC AAM84513;

DT 07-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen SEQ ID NO:12106.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KM Cytostatic; gene therapy; vaccine; metastasis.

XX Homo sapiens.

OS MO200157182-A2.

PN 09-AUG-2001.

PD 17-JAN-2001; 2001MO-US01354.

PF 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 01-NOV-2000; 2000US-0241826.

PR 08-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246539.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251899.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX  
XX MPI: 2001-483426/52.  
DR N-PSDB; AAK57294.  
PT  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX  
PS Claim 11; SEQ ID NO 12106; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX  
SQ Sequence 237 AA;  
Query Match 2.7%; Score 7; DB 22; Length 237;  
Best Local Similarity 100.0%; Pred. No. 4.9e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ID AAU17162 standard; Protein; 237 AA.  
XX  
XX AAU17162;  
AC  
XX  
XX  
DT 07-NOV-2001 (first entry)  
XX  
XX  
DE Novel signal transduction pathway protein, Seq ID 727.  
XX  
XX  
XX Neutropenic; cytostatic; dermatological; immunosuppressive; tumour;  
XX antineoplastic; anti-HIV; antibacterial; anti-inflammatory; cancer;  
XX immune system disorder; rheumatoid arthritis; inflammatory condition;  
XX organ transplant rejection; infection; hepatitis C; blood disorder;  
XX sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
XX chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
XX cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
XX reproductive system; gastrointestinal; liver disorder; AIDS;  
XX acquired immune deficiency syndrome.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200154733-A1.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01312.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.

RESULT 51  
AAU17162

PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-465460/50.  
 DR N-PSDB; AAS27079.  
 XX  
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
 PT proposing disorders related to the proteins, including cancers, immune  
 PT disorders and neuronal disorders -  
 XX  
 PS Claim 1; SEQ ID NO 727; 860pp; English.  
 XX  
 CC The invention relates to novel isolated polypeptides (I), and  
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
 CC diagnosing, preventing and treating diseases including immune system  
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
 CC transplant rejections and graft versus host disease, infectious diseases  
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
 CC other blood-related disorders (sickle cell anemia), myeloproliferative  
 CC disorders, primary haematopoietic disorders, hyperproliferative  
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal  
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
 CC Addison's disease), reproductive system disorders, gastrointestinal  
 CC disorders (inflammatory disorders), liver disorders (cirrhosis),  
 CC as stimulators of B-cell responsiveness to pathogens, activators of  
 CC T-cells, to induce higher affinity antibodies, and as a means to induce  
 CC tumour proliferation in pathologies e.g. acquired immune deficiency  
 CC syndrome (AIDS). AAU17059-AAU1763 represent novel signal transduction  
 CC pathway protein, amino acid sequences of the invention.  
 XX

Query Match	2.7%; Score 7; DB 22; Length 237;
Best Local Similarity	100.0%; Pred. No. 4.9e+02;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	149 VIGISOR 155 
Db	215 VIGISOR 221

RESULT 52  
 ID AAU34184 standard; Protein; 258 AA.  
 XX AAU34184;  
 AC  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX

Staphylococcus aureus cellular proliferation protein #460.

XX Antisense; prokaryotic cellular proliferation protein;  
KM antibiotic; antibacterial; drug design.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO200170955-A2.  
XX  
XX  
XX 27-SEP-2001.  
XX  
XX 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
XX 23-MAY-2000; 2000US-206848P.  
XX 26-MAY-2000; 2000US-207727P.  
XX 23-OCT-2000; 2000US-242578P.  
XX 27-NOV-2000; 2000US-253625P.  
XX 22-DEC-2000; 2000US-257931P.  
XX 16-FEB-2001; 2001US-269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
XX N-PSDB; AAS52043.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Example 3; Seq ID No 5680; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 258 AA;  
XX  
XX Query Match 2.7%; Score 7; DB 22; Length 258;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 147 QPVIGIS 153  
DB 56 QPVIGIS 62

RESULT 53  
AAM41810  
ID AAM41810 standard; Protein; 272 AA.  
XX  
AC AAM41810;  
XX  
XX 22-OCT-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 6741.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KM peripheral nervous system; neuropathy; central nervous system; CNS;  
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.  
XX 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac R;  
XX  
XX WPI; 2001-442253/47.  
XX N-PSDB; AAI60966.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Example 2; SEQ ID NO 6741; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cyrostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localized neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 272 AA;  
XX  
XX Query Match 2.7%; Score 7; DB 22; Length 272;  
Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 143 AAAYQPV 149  
DB 73 AAAYQPV 79

RESULT 54  
AAM41811  
ID AAM41811 standard; Protein; 272 AA.  
XX  
AC AAM41811;  
XX  
XX



DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 6742.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemokine;  
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Dymanc RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AAI60967.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 2; SEQ ID NO 6742; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 272 AA;  
 XX  
 Query Match 2.7%; Score 7; DB 22; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 143 AAAVQPV 149  
 Db 73 AAAVQPV 79  
 XX  
 RESULT 55  
 ABB59873  
 ID ABB59873 standard; Protein; 282 AA.

XX  
 AC ABB59873;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 6411.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB; ABL03976.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 6411; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABU16176-ABU16175), expressed DNA  
 CC sequences (ABU16176-ABU16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 282 AA;  
 XX  
 Query Match 2.7%; Score 7; DB 22; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 5.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 LNWIDAQ 53  
 Db 178 LNWIDAQ 184  
 XX  
 RESULT 56  
 ABG17587  
 ID ABG17587 standard; Protein; 296 AA.  
 XX  
 AC ABG17587;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #17578.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.

```
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
FA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR
XX N-PSDB; AAS81774.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS
XX Claim 20; SEQ ID No 47946; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences, (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 296 AA;
SQ
XX
XX Query Match 2.7%; Score 7; DB 22; Length 296;
XX Best Local Similarity 100.0%; Pred. No. 5.9e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 105 TSOALPA 111
Db 9 TSOALPA 15
XX
XX RESULT 57
XX AAU36580
XX ID AAU36580 standard; Protein; 300 AA.
XX
XX AAU36580;
AC
XX
XX 14-FEB-2002 (first entry)
DT
XX
XX Staphylococcus aureus cellular proliferation protein #750.
DE
XX
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
OS
XX
XX WO200170955-A2.
PN
XX
XX 27-SEP-2001.
PD
```

```
XX 21-MAR-2001; 2001WO-US09180.
PF
XX
XX 21-MAR-2000; 2000US-191078P.
PR
XX 23-MAY-2000; 2000US-206848P.
PR
XX 26-MAY-2000; 2000US-207727P.
PR
XX 23-OCT-2000; 2000US-242578P.
PR
XX 27-NOV-2000; 2000US-253625P.
PR
XX 22-DEC-2000; 2000US-257931P.
PR
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haeselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
PI
XX WPI: 2001-611495/70.
DR
XX N-PSDB; AAS5439.
DR
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PT
XX
XX Example 3; Seq ID No 12173; 51pp; English.
PS
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 300 AA;
SQ
XX
XX Query Match 2.7%; Score 7; DB 22; Length 300;
XX Best Local Similarity 100.0%; Pred. No. 6e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 147 QPVIGIS 153
Db 73 QPVIGIS 79
XX
XX RESULT 58
XX ABB68531
XX ID ABB68531 standard; Protein; 302 AA.
XX
XX ABB68531;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 32385.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
```

```

XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656660/75.
DR
XX N-PSDB; ABL12634.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX
XX Disclosure; SEQ ID NO 32385; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB57737-AB372072).
XX (AB57737-AB372072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 302 AA:
SQ
XX
XX Query Match 2.7%; Score 7; DB 22; Length 302;
XX Best Local Similarity 100.0%; Pred. No. 6e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 85 SGEAGVP 91
XX |||||
XX 176 SGEAGVP 182
XX
XX RESULT 59
XX AAY06112
XX ID AAY06112 standard; Protein; 323 AA.
XX
XX AAY06112;
XX
XX 16-AUG-1999 (first entry)
XX
XX Nelson Bay virus sigma 3 protein.
XX
XX Orthoreovirus; reovirus; NBV; sigma 3; viral cell attachment;
XX cell fusion.
XX
XX Nelson Bay virus.
XX
XX WO9924582-A1.
XX
XX 20-MAY-1999.
XX
XX 06-NOV-1998; 98WO-CA01046.
XX
XX 07-NOV-1997; 97US-0965708.
XX
XX (UYDA-) UNIV DALHOUSIE.
XX
XX Duncan R;
XX
XX WPI; 1999-327410/27.
XX
XX N-PSDB; AAX58669.
XX
XX Fusogenic proteins from non-enveloped reoviruses

```

```

XX Example 6; Page 51-52; 57pp; English.
PS
XX
XX The present sequence represents sigma 3 protein, the viral cell
XX attachment protein, of Nelson Bay virus (NBV). The coding region
XX for this protein lies downstream of the fusogenic P1 protein
XX coding region in the NBV genome. 2 Unrelated fusion proteins
XX responsible for cell-cell fusion induced by avian reovirus (ARV)
XX and the only 2 fusogenic mammalian reoviruses, NBV and baboon
XX reovirus (BRV), are identified in this invention. These proteins
XX are named P11 for ARV and NBV (see AAY06104, AAY06107 and AAY06110) and
XX P15 for BRV (see AAY06113). Fusogenic P11 and P15 proteins are
XX useful for promoting cell and liposome fusion, e.g. for production
XX of hybridomas and for liposome-mediated delivery of bioactive
XX agents to cells.
XX
XX Sequence 323 AA:
SQ
XX
XX Query Match 2.7%; Score 7; DB 20; Length 323;
XX Best Local Similarity 100.0%; Pred. No. 6.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 52 AQSGLAS 58
XX |||||
XX 118 AQSGLAS 124
XX
XX RESULT 60
XX AAB41611
XX ID AAB41611 standard; Protein; 332 AA.
XX
XX AAB41611;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF1375 polypeptide sequence SEQ ID NO:2750.
XX
XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
XX vulnerrary; antipsoiatric; antiparkinsonian; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antitumetic; antihydroid;
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX
XX 02-APR-1999; 99US-0127636.
XX
XX 05-APR-1999; 99US-0127728.
XX
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX
XX N-PSDB; AAC75820.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX

```



```

PI Yamamoto RT, Xu HH;
XX
DR MPI: 2001-611495/70.
DR N-PSDB; AAB53331.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 11065; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 335 AA;

Query Match 2.7%; Score 7; DB 22; Length 335;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 IAIGAGI 188
DB 198 IAIGAGI 204

RESULT 63
AAB90557
ID ABB90557 standard; Protein; 356 AA.
XX
AC ABB90557;
XX
D7 29-JUL-2002 (first entry)
XX
DE Chlamydia pneumoniae cp7106 protein, SEQ ID NO:63.
XX
KW Chlamydial infection; antigen; immunogen; vaccine; diagnosis;
KW human respiratory disease; cardiovascular disease; atherosclerosis;
KW coronary artery disease; carotid artery stenosis; myocardial infarction;
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW strain CWL029.
XX
XX Chlamydia pneumoniae.
XX
OS
XX
PN WO200202606-A2.
XX
PD 10-JAN-2002.
XX
PF 03-JUL-2001; 2001WO-IB01445.
XX
XX 03-JUL-2000; 2000GB-0016363.
PR 11-JUL-2000; 2000GB-0017047.
PR 21-JUL-2000; 2000GB-0017983.
PR 07-AUG-2000; 2000GB-0019368.
PR 18-AUG-2000; 2000GB-0020440.
PR 14-SEP-2000; 2000GB-0022583.
PR 10-NOV-2000; 2000GB-0027549.
PR 22-DEC-2000; 2000GB-0031706.

```

```

XX
PA (CHIR-) CHIRON SPA.
XX
XX Ratti G, Grandi G;
XX
DR MPI: 2002-154726/20.
DR N-PSDB; ABL91215.
XX
XX Novel Chlamydia pneumoniae protein useful in the manufacture of a
PT medicament for treatment or prevention of infection due to Chlamydia,
PT preferably Chlamydia pneumoniae, and for diagnostic purposes -
XX
PS Claim 1; Page 73; 364pp; English.
XX
XX Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
CC them. The proteins are predicted to be immunogenic and may therefore be
CC useful in vaccine production and for diagnostic purposes. Chlamydia
CC pneumoniae is a common cause of respiratory disease in humans, and is
CC also involved in the development of cardiovascular diseases such as
CC atherosclerosis, coronary artery disease, carotid artery stenosis,
CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
CC claudication and stroke. The proteins and nucleic acids of the invention
CC may be used in vaccines and pharmaceutical compositions for the
CC prevention or treatment of chlamydial infections, particularly Chlamydia
CC pneumoniae infections. The proteins may also be used in the detection of
CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC DNA probe assay or blotting techniques for determining Chlamydia
CC pneumoniae gene expression. The present sequence represents a
CC specifically claimed Chlamydia pneumoniae protein of the invention.
XX
SQ Sequence 356 AA;

Query Match 2.7%; Score 7; DB 23; Length 356;
Best Local Similarity 100.0%; Pred. No. 6.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDGLGTG 170
DB 2 KDGLGTG 8

RESULT 64
AAB01333
ID AAB01333 standard; Protein; 357 AA.
XX
AC AAB01333;
XX
D7 25-SEP-2000 (first entry)
XX
DE FraZ1 polypeptide of corn.
XX
XX Plasmid; transformation; gene expression; hybrid; sunflower;
KW safflower; cotton; soybean; peanut; coconut; oil palm; corn;
KW vegetable oil; homoplasmic plant.
XX
XX Zea mays.
XX
OS
XX
PN Key Location/Qualifiers
XX
FT Misc-difference 270 /note= "Unidentified amino acid"
FT Misc-difference 280 /note= "Unidentified amino acid"
FT Misc-difference 307 /note= "Unidentified amino acid"
FT Misc-difference 314 /note= "Unidentified amino acid"
FT Misc-difference 330 /note= "Unidentified amino acid"
FT Misc-difference 332 /note= "Unidentified amino acid"
FT Misc-difference 334 /note= "Unidentified amino acid"
FT /note= "Unidentified amino acid"

```

FT Misc-difference 343 /note= "Unidentified amino acid"  
 FT Misc-difference 346 /note= "Unidentified amino acid"  
 FT  
 XX WO200032799-A1.  
 XX  
 XX 08-JUN-2000.  
 XX  
 XX 24-NOV-1999; 99WO-US28103.  
 XX  
 XX 25-NOV-1998; 98US-0109892.  
 XX  
 XX (CALJ ) CALGENE LLC.  
 XX  
 XX Chaudhuri S;  
 XX  
 XX WPI; 2000-412344/35.  
 XX  
 XX  
 XX Transforming plant cell plastids useful in plant breeding involves  
 PT introducing a construct comprising functional promoter in and a DNA  
 PT sequence of interest into a plant cell having altered plastid  
 PT morphology  
 PT  
 XX  
 XX Disclosure; Page 69-70; 73pp; English.  
 XX  
 XX The FtsZ sequences of plants have been shown to be a component of  
 CC the plastid division mechanism. Disruption of this gene can result  
 CC in mutant cell lines having altered numbers and/or size of plastids.  
 CC It is therefore possible to engineer plant cells comprising 20,000  
 CC plant cell plastid (PCP) involves introducing a construct, comprising  
 CC a promoter functional in a PCP operably associated with the DNA  
 CC sequence of interest into a PC having an altered plant plastid  
 CC morphology, selected from altered plastid size and plastid number.  
 CC The expression of FtsZ nucleic acid sequences in a plant cell  
 CC produces an altered number and/or size of the chloroplasts within the  
 CC host cell. The plants can be used as targets for plastid  
 CC transformation. Plant expression or transcription constructs having a  
 CC plant FtsZ as the DNA sequence of interest can be used in the  
 CC production of vegetable oils for edible and industrial uses, using  
 CC temperate oilseed crops such as rapeseed, sunflower, safflower,  
 CC cotton, soybean, peanut, coconut and oil palms, and corn. Using tissue  
 CC derived from the plants in the plastid transformation methods,  
 CC increases the efficiency of foreign DNA transformation into the plant  
 CC cell plastid. The plant tissue source used in the transformation  
 CC methods contains larger plastids, which provides for a larger surface  
 CC area for the DNA to penetrate the plastid membrane during  
 CC transformation. The method provides for a more efficient approach to  
 CC obtaining homoplasmic plants.  
 CC  
 XX  
 SQ Sequence 357 AA;  
 Query Match 2.7%; Score 7; DB 21; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 AAAQPV 149  
 Db 252 AAAQPV 258

RESULT 65  
 AAUS6062  
 ID AAUS6062 standard; Protein; 363 AA.  
 XX  
 XX AAUS6062;  
 XX  
 XX 27-FEB-2002 (first entry)  
 XX  
 XX Propionibacterium acnes immunogenic protein #16958.  
 DE  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 XX Propionibacterium acnes.  
 OS  
 XX WO200181581-A2.  
 XX  
 XX  
 XX 01-NOV-2001.  
 XX  
 XX 20-APR-2001; 2001WO-US12865.  
 XX  
 XX 21-APR-2000; 2000US-199047P.  
 XX  
 XX 02-JUN-2000; 2000US-208841P.  
 XX  
 XX 07-JUL-2000; 2000US-216747P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 XX I'matsonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 XX WPI; 2001-616774/71.  
 XX  
 XX N-PSDB; AAS59573.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris .  
 PT  
 XX  
 XX Example 1; SEQ ID No 17257; 1069pp; English.  
 PS  
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 363 AA;  
 Query Match 2.7%; Score 7; DB 22; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 EASEGPG 124  
 Db 333 EASEGPG 339

RESULT 66  
 ABB90566  
 ID ABB90566 standard; Protein; 365 AA.  
 XX  
 XX ABB90566;  
 XX  
 XX 29-JUL-2002 (first entry)  
 XX  
 XX Chlamydia pneumoniae cp7108 protein, SEQ ID NO:81.  
 DE  
 XX Chlamydial infection; antigen; immunogen; vaccine; diagnosis;

KW human respiratory disease; cardiovascular disease; atherosclerosis;  
KW coronary artery disease; carotid artery stenosis; myocardial infarction;  
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;  
KW strain CML029.  
XX  
OS Chlamydia pneumoniae.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..23 /label= Signal\_peptide  
FT Protein 24..365 /note= "Mature protein"  
XX  
PN WO200202606-A2.  
XX  
PD 10-JAN-2002.  
XX  
PF 03-JUL-2001; 2001WO-1B01445.  
XX  
PR 03-JUL-2000; 2000GB-0016363.  
PR 11-JUL-2000; 2000GB-0017047.  
PR 21-JUL-2000; 2000GB-0017983.  
PR 07-AUG-2000; 2000GB-0019368.  
PR 18-AUG-2000; 2000GB-0020440.  
PR 14-SEP-2000; 2000GB-0022583.  
PR 10-NOV-2000; 2000GB-0027549.  
PR 22-DEC-2000; 2000GB-0031706.  
XX  
PA (CHIR-) CHIRON SPA.  
XX  
PI Ratti G, Grandi G;  
XX  
DR WPI; 2002-154726/20.  
DR N-PSDB; ABL91224.  
XX  
PT Novel Chlamydia pneumoniae protein useful in the manufacture of a  
PT medicament for treatment or prevention of infection due to Chlamydia,  
PT preferably Chlamydia pneumoniae, and for diagnostic purposes -  
XX  
PS Claim 1; Page 79-80; 364pp; English.  
XX  
CC Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia  
CC pneumoniae (strain CML029), and ABL91184-ABL91373 represent DNA encoding  
CC them. The proteins are predicted to be immunogenic and may therefore be  
CC useful in vaccine production and for diagnostic purposes. Chlamydia  
CC pneumoniae is a common cause of respiratory disease in humans, and is  
CC also involved in the development of cardiovascular diseases such as  
CC atherosclerosis, coronary artery disease, carotid artery stenosis,  
CC myocardial infarction, cerebrovascular disease, aortic aneurysm,  
CC claudication and stroke. The proteins and nucleic acids of the invention  
CC may be used in vaccines and pharmaceutical compositions for the  
CC prevention or treatment of chlamydial infections, particularly Chlamydia  
CC pneumoniae infections. The proteins may also be used in the detection of  
CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched  
CC DNA probe assay or blotting techniques for determining Chlamydia  
CC pneumoniae gene expression. The present sequence represents a  
CC specifically claimed Chlamydia pneumoniae protein of the invention.  
XX  
SQ Sequence 365 AA;  
Query Match 2.7%; Score 7; DB 23; Length 365;  
Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 164 KDLGTG 170  
Db 142 KDLGTG 148  
RESULT 67  
AAU87261  
ID AAU87261 standard; Protein; 396 AA.  
XX

AC AAU87261;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Novel central nervous system protein #171.  
XX  
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
KW myocardial infarction; wound healing; cell proliferation; skin aging;  
KW food additive; food preservative; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200155318-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01332.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232387.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 06-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM,  
PI  
XX WPI; 2001-581633/65.  
XX  
DR N-PSDB; ABK43591.  
DR  
XX  
PI New isolated nucleic acid encoding a protein for diagnosing,  
PT preventing, treating or ameliorating medical conditions and used as  
PT food additives or preservatives -  
XX  
XX Claim 9; SEQ ID No 779; 837pp; English.  
XX  
CC The invention describes an isolated nucleic acid molecule (I) encoding a  
CC novel central nervous system protein. (II) and polypeptides (III) encoded  
CC by (I), are used to treat a medical conditions and in diagnosis of a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
CC angiotensis, nervous system disorders e.g. Alzheimer's disease and  
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
CC adenocarcinomas and irritable bowel syndrome, reproductive system  
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
CC acute kidney failure and blood related disorders e.g. myocardial  
CC infarction. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
Query Match 2.7%; Score 7; DB 22; Length 396;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 143 AAAVQPV 149  
Db 234 AAAVQPV 240  
RESULT 68  
AAU17285  
ID AAU17285 standard; Protein; 399 AA.  
XX  
XX AAU17285;  
AC  
DT 07-NOV-2001 (first entry)  
XX  
XX Novel signal transduction pathway protein, Seq ID 850.  
DE  
XX  
KM Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;



KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS;  
KW acquired immune deficiency syndrome.  
XX Homo sapiens.  
XX OS  
XX PN WO200154733-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01312.  
XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227709.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0233081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0238935.  
PR 13-OCT-2000; 2000US-0238937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-465460/50.  
DR N-PSDB; AAS27202.  
XX  
PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
PT prognosing disorders related to the proteins, including cancers, immune  
PT disorders and neuronal disorders -  
XX  
PS Claim 1; SEQ ID No 850; 880bp; English.  
XX  
CC The invention relates to novel isolated polypeptides (I), and  
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
CC diagnosing, preventing and treating diseases including immune system  
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
CC transplant rejections and graft versus host disease, infectious diseases  
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
CC other blood-related disorders (sickle cell anemia), myeloproliferative  
CC disorders, primary haematopoietic disorders, hyperproliferative  
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
CC disorders (e.g. Alzheimer's disease and Parkinson's disease), chromosomal  
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal  
CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
CC Addison's disease), reproductive system disorders, gastrointestinal  
CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
CC as stimulators of B-cell responsiveness to pathogens, activators of  
CC T-cells, to induce higher affinity antibodies, and as a means to induce  
CC tumour proliferation in pathologies e.g. acquired immune deficiency  
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction  
CC pathway protein, amino acid sequences of the invention.  
XX  
Query Match 2.7%; Score 7; DB 22; Length 359;  
Best Local Similarity 100.0%; Pred. No. 7.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 143 AAAYQPV 149  
| | | | |  
Db 234 AAAYQPV 240  
RESULT 69  
ABG03405  
ID ABG03405 standard; Protein; 413 AA.  
XX  
AC ABG03405;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #3396.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.

XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS67592.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 33764; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 413 AA;  
Query Match 2.7%; Score 7; DB 22; Length 413;  
Best Local Similarity 100.0%; Pred. No. 7.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 114 TEIQEAS 120  
| | | | |  
Db 104 TEIQEAS 110  
RESULT 70  
ABG16425  
ID ABG16425 standard; Protein; 413 AA.  
XX  
AC ABG16425;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #16416.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.

```

XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX {HYSE-} HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS80612.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 46784; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 413 AA;

Query Match 2.7%; Score 7; DB 22; Length 413;
Best Local Similarity 100.0%; Pred. No. 7.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 TEIOEAS 120
   |||||
Db 104 TEIOEAS 110

RESULT 71
AAU38368
ID AAU38368 standard; Protein; 465 AA.
XX
AC AAU38368;
XX
DT 14-FEB-2002 (first entry)
XX
DE Salmonella typhi cellular proliferation protein #259.
XX
KM Antisense; prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.

```

```

PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS56227.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 13961; 511bp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 465 AA;

Query Match 2.7%; Score 7; DB 22; Length 465;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 GTLGYYL 173
   |||||
Db 190 GTLGYYL 196

RESULT 72
ABP72633
ID ABP72633 standard; Protein; 483 AA.
XX
AC ABP72633;
XX
DT 11-JUN-2003 (first entry)
XX
DE Chelonius sp. chitinase.
XX
KM Insecticide; pesticide; insect control; insect; toxin; chitinase;
KM enzyme; neuropeptide; transgenic plant; crop protection.
XX
OS Chelonius sp.
XX
XX Key Location/Qualifiers
FH 1..17
FT Peptide /label= Signal_peptide
FT Protein /label= Signal_peptide
FT Active-site /label= Mature_protein
FT 138..145
XX

```

```

PN WO2003014150-A2.
XX
XX 20-FEB-2003.
PD
XX 06-AUG-2002; 2002WO-GB03598.
XX
XX 08-AUG-2001; 2001GB-0019274.
PR
XX (UYDU-) UNIV DURHAM.
XX (ENVI-) DEPT ENVIRONMENT FOOD & RURAL AFFAIRS.
PA
XX Gatehouse JA, Fitches EC, Edwards JP;
PI WPI; 2003-278469/27.
XX
XX Fusion protein useful for combating insect pests, comprises a
PT translocating moiety comprising a plant protein capable of acting as a
PT carrier to translocate toxic moiety inside plant pathogen, and a toxic
PT moiety -
XX
XX Claim 7; Fig 14; 51pp; English.
XX
XX The present sequence is that of Chelonus sp. chitinase. This
CC protein is used in claimed fusion proteins of the invention. Such
CC fusion proteins comprise a translocating moiety and a toxic
CC moiety, where the translocating moiety is a plant protein (e.g. a
CC lectin) capable of acting as a carrier to translocate the toxic
CC moiety across the gut wall of a plant pathogen, and the toxic
CC moiety is an insect-derived peptide or protein capable of causing
CC deleterious effects on growth, development, reproduction or
CC mortality in pest insects. Suitable insect peptides and proteins
CC include allatotransins, chitinase, diuretic hormone and their
CC metabolites and analogues. Polynucleotides encoding the fusion
CC protein, vectors, host cells and transgenic plants that are
CC resistant to disease are also provided. The fusion protein is
CC target-specific, and resists degradation in the insect gut.
CC
XX
XX Sequence 483 AA;
SQ
XX
XX Query Match 2.7%; Score 7; DB 24; Length 483;
XX Best Local Similarity 100.0%; Pred. No. 9e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 59 APVSGAG 65
XX |||||
Db 281 APVSGAG 287
XX
XX RESULT 73
XX ABR40757
XX ID ABR40757 standard; Protein; 484 AA.
XX
XX ABR40757;
XX
XX 16-MAY-2003 (first entry)
XX
XX Zea mays oil trait related protein sequence SEQ ID NO:334.
XX
XX Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;
XX receptor-like protein kinase; mitogen activated protein kinase; oil;
XX LIP5-like transcription factor caleosin; ATP citrate lyase; SNF1;
XX CKC-like transcription factor; antisense inhibition; co-suppression;
XX transgenic plant.
XX
XX Zea mays.
XX
XX WO2003002751-A2.
XX
XX 09-JAN-2003.
XX
XX 27-JUN-2002; 2002WO-US20152.
XX
XX 29-JUN-2001; 2001US-301913P.
XX

```

```

XX
XX (DUBO ) DU PONT DE NEMOURS & CO E I.
PA
XX (PION-) PIONEER HI-BRED INT INC.
PA
XX Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvey LT;
PI Jones TV, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H;
PI Shen B, Tarczynski MC;
XX
XX WPI; 2003-201509/19.
XX
XX N-PSDB; ACC00792.
XX
XX Novel nucleotide fragment encoding polypeptides having receptor-like
PT protein kinase activity, caleosin-like activity, useful for altering
PT oil phenotypes in plants such as sunflower, coconut, soybean, wheat and
PT rice -
XX
XX Claim 12; Page 352-353; 542pp; English.
XX
XX The present invention describes an isolated nucleotide fragment (I)
CC comprising a nucleic acid sequence (NS) chosen from a NS encoding a
CC polypeptide (PP) having receptor-like protein kinase activity, mitogen
CC activated protein (MAP)-kinase activity, LIP5-like transcription factor
CC activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like
CC activity and CKC-like transcription factor activity. Also described:
CC (1) complement (II) of (I); (2) a chimeric construct (III) comprising (I)
CC or (II), operably linked to a regulatory sequence; (3) a plant (IV)
CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and
CC (5) oil obtained from (V). (I) or its part can be used in antisense
CC inhibition or co-suppression in a transformed plant. (III) is useful for
CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,
CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for
CC creating transgenic plants having altered lipid profiles. (I) can also
CC be used as a hybridisation probe. ACC00656 to ACC00688 and ABR40591 to
CC ABR40879 represent sequences used in the exemplification of the present
CC invention.
XX
XX Sequence 484 AA;
SQ
XX
XX Query Match 2.7%; Score 7; DB 24; Length 484;
XX Best Local Similarity 100.0%; Pred. No. 9.1e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 EAYSGG 24
XX |||||
Db 384 EAYSGG 390
XX
XX RESULT 74
XX ABR40829
XX ID ABR40829 standard; Protein; 485 AA.
XX
XX ABR40829;
XX
XX 16-MAY-2003 (first entry)
XX
XX Zea mays oil trait related protein sequence SEQ ID NO:421.
XX
XX Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;
XX receptor-like protein kinase; mitogen activated protein kinase; oil;
XX LIP5-like transcription factor caleosin; ATP citrate lyase; SNF1;
XX CKC-like transcription factor; antisense inhibition; co-suppression;
XX transgenic plant.
XX
XX Zea mays.
XX
XX WO2003002751-A2.
XX
XX 09-JAN-2003.
XX
XX 27-JUN-2002; 2002WO-US20152.
XX
XX 29-JUN-2001; 2001US-301913P.
XX

```

PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Allen SM, Allen WB, Cahoon RE, Epelbaum S, Farnodu OO, Harvell LT;  
 PI Jones TU, Kinney AJ, Kleim TM, Li C, Oliveira IC, Sakai H;  
 PI Shen B, Tarczyński MC;  
 XX  
 DR WPI, 2003-201509/19.  
 XX  
 PT Novel nucleotide fragment encoding polypeptides having receptor-like  
 PT protein kinase activity, caleosin-like activity, useful for altering  
 PT oil phenotypes in plants such as sunflower, coconut, soybean, wheat and  
 PT rice -  
 XX  
 PS Claim 12; Page 458-459; 542pp; English.  
 XX  
 CC The present invention describes an isolated nucleotide fragment (I)  
 CC comprising a nucleic acid sequence (NS) chosen from a NS encoding a  
 CC polypeptide (PP) having receptor-like protein kinase activity, mitogen  
 CC activated protein (MAP)-kinase activity, LIP15-like transcription factor  
 CC activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like  
 CC activity and CKC-like transcription factor activity. Also described:  
 CC (1) complement (II) of (I); (2) a chimeric construct (III) comprising (I)  
 CC or (II), operably linked to a regulatory sequence; (3) a plant (IV)  
 CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and  
 CC (5) oil obtained from (V). (I) or its part can be used in antisense  
 CC inhibition or co-suppression in a transformed plant. (III) is useful for  
 CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,  
 CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for  
 CC creating transgenic plants having altered lipid profiles. (I) can also  
 CC be used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to  
 CC ABR40879 represent sequences used in the exemplification of the present  
 CC invention.  
 CC  
 XX  
 SQ Sequence 485 AA;  
 OY 18 EAYGSGG 24  
 DB 385 EAYGSGG 391  
 Query Match 2.7%; Score 7; DB 24; Length 485;  
 Best Local Similarity 100.0%; Pred. No. 9.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 75  
 ABR40834  
 ID ABR40834 standard; Protein; 485 AA.  
 AC ABR40834;  
 XX  
 DT 16-MAY-2003 (first entry)  
 XX  
 DE Zea mays oil trait related protein sequence SEQ ID NO:426.  
 XX  
 KW Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;  
 KW receptor-like protein kinase; mitogen activated protein kinase; oil;  
 KW LIP15-like transcription factor; caleosin; ATP citrate lyase; SNF1;  
 KW CKC-like transcription factor; antisense inhibition; co-suppression;  
 KW transgenic plant.  
 XX  
 OS Zea mays.  
 OS  
 XX  
 PN WO2003002751-A2.  
 XX  
 PD 09-JAN-2003.  
 XX  
 PF 27-JUN-2002; 2002WO-US20152.  
 PF  
 XX 29-JUN-2001; 2001US-301913P.  
 FR  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA (PION-) PIONEER HI-BRED INT INC.

XX  
 PI Allen SM, Allen WB, Cahoon RE, Epelbaum S, Farnodu OO, Harvell LT;  
 PI Jones TU, Kinney AJ, Kleim TM, Li C, Oliveira IC, Sakai H;  
 PI Shen B, Tarczyński MC;  
 XX  
 DR WPI, 2003-201509/19.  
 XX  
 PT Novel nucleotide fragment encoding polypeptides having receptor-like  
 PT protein kinase activity, caleosin-like activity, useful for altering  
 PT oil phenotypes in plants such as sunflower, coconut, soybean, wheat and  
 PT rice -  
 XX  
 PS Claim 12; Page 466-468; 542pp; English.  
 XX  
 CC The present invention describes an isolated nucleotide fragment (I)  
 CC comprising a nucleic acid sequence (NS) chosen from a NS encoding a  
 CC polypeptide (PP) having receptor-like protein kinase activity, mitogen  
 CC activated protein (MAP)-kinase activity, LIP15-like transcription factor  
 CC activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like  
 CC activity and CKC-like transcription factor activity. Also described:  
 CC (1) complement (II) of (I); (2) a chimeric construct (III) comprising (I)  
 CC or (II), operably linked to a regulatory sequence; (3) a plant (IV)  
 CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and  
 CC (5) oil obtained from (V). (I) or its part can be used in antisense  
 CC inhibition or co-suppression in a transformed plant. (III) is useful for  
 CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,  
 CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for  
 CC creating transgenic plants having altered lipid profiles. (I) can also  
 CC be used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to  
 CC ABR40879 represent sequences used in the exemplification of the present  
 CC invention.  
 CC  
 XX  
 SQ Sequence 485 AA;  
 OY 18 EAYGSGG 24  
 DB 385 EAYGSGG 391  
 Query Match 2.7%; Score 7; DB 24; Length 485;  
 Best Local Similarity 100.0%; Pred. No. 9.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Search completed: November 25, 2003, 13:39:43  
 Job time : 44 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:41:15 ; Search time 31 Seconds

(without alignments)  
1564.793 Million cell updates/sec

Title: US-10-057-951-2

Sequence: 1 MLAWVQAFVLSNMDLAEX.....PVDPOEGSTPLMGAGTPGA 263

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 673684 seqs, 184443283 residues

Word size : 0

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PC1US\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	263	100.0	263	9 US-09-084-491A-2	Sequence 2, Appl1
2	263	100.0	263	14 US-10-102-704-2	Sequence 2, Appl1
3	263	100.0	263	12 US-10-057-951-2	Sequence 2, Appl1
4	250	95.1	263	12 US-10-210-951-44	Sequence 44, Appl1
5	250	95.1	263	12 US-10-211-884-44	Sequence 44, Appl1
6	66	25.1	66	9 US-09-864-761-38457	Sequence 38457, A
7	8	3.0	457	15 US-10-156-761-10667	Sequence 10667, A
8	8	3.0	716	12 US-09-981-151A-55	Sequence 55, Appl1
9	7	2.7	43	9 US-09-864-761-45155	Sequence 45155, A
10	7	2.7	69	12 US-10-029-386-27831	Sequence 27831, A
11	7	2.7	11	12 US-10-279-579A-18	Sequence 18, Appl1
12	7	2.7	154	10 US-09-738-626-3788	Sequence 3788, Ap
13	7	2.7	237	10 US-09-764-868-727	Sequence 727, App
14	7	2.7	258	9 US-09-815-242-5680	Sequence 5680, Ap
15	7	2.7	300	9 US-09-815-242-12173	Sequence 12173, A

16	7	2.7	323	9 US-09-943-002-12	Sequence 12, Appl1
17	7	2.7	335	9 US-09-815-242-11065	Sequence 11065, A
18	7	2.7	399	10 US-09-764-868-850	Sequence 850, App
19	7	2.7	465	9 US-09-815-242-13961	Sequence 13961, A
20	7	2.7	545	15 US-10-156-761-11459	Sequence 11459, A
21	7	2.7	562	10 US-09-738-626-5317	Sequence 5317, Ap
22	7	2.7	684	10 US-09-946-678-2	Sequence 2, Appl1
23	7	2.7	701	9 US-09-815-242-13411	Sequence 13411, A
24	7	2.7	706	10 US-09-738-626-6019	Sequence 6019, Ap
25	7	2.7	937	10 US-09-974-298-129	Sequence 129, Appl
26	7	2.7	1156	15 US-10-099-285-72	Sequence 72, Appl
27	7	2.7	2472	7 US-09-815-242-5064	Sequence 5064, Ap
28	6	2.3	7	11 US-09-792-286-183	Sequence 183, App
29	6	2.3	7	11 US-09-792-286-183	Sequence 183, App
30	6	2.3	16	15 US-10-206-639-74	Sequence 74, Appl1
31	6	2.3	20	12 US-10-032-221B-28	Sequence 28, Appl1
32	6	2.3	27	9 US-09-864-761-35595	Sequence 35595, A
33	6	2.3	32	9 US-09-864-761-48415	Sequence 48415, A
34	6	2.3	34	12 US-10-160-162-121	Sequence 121, App
35	6	2.3	34	12 US-09-820-649-121	Sequence 121, App
36	6	2.3	37	15 US-10-106-698-8286	Sequence 8286, Ap
37	6	2.3	38	9 US-09-864-761-38182	Sequence 38182, A
38	6	2.3	39	11 US-09-983-802-589	Sequence 589, App
39	6	2.3	42	9 US-09-864-761-34850	Sequence 34850, A
40	6	2.3	46	15 US-10-106-638-8514	Sequence 8514, Ap
41	6	2.3	50	9 US-09-864-761-40317	Sequence 40317, A
42	6	2.3	51	9 US-09-864-761-37671	Sequence 37671, A
43	6	2.3	52	12 US-10-029-386-29128	Sequence 29128, A
44	6	2.3	52	16 US-10-218-102-371	Sequence 371, App
45	6	2.3	53	11 US-09-983-802-588	Sequence 588, App
46	6	2.3	56	12 US-10-029-386-31083	Sequence 31083, A
47	6	2.3	58	15 US-10-007-2804-184	Sequence 184, App
48	6	2.3	59	9 US-09-879-957-215	Sequence 215, App
49	6	2.3	60	9 US-09-864-761-47913	Sequence 47913, A
50	6	2.3	60	15 US-10-156-761-7793	Sequence 7793, Ap
51	6	2.3	61	10 US-09-912-787-77	Sequence 77, Appl1
52	6	2.3	61	14 US-10-014-326-63	Sequence 63, Appl1
53	6	2.3	62	9 US-09-864-761-40363	Sequence 40363, A
54	6	2.3	63	10 US-09-738-626-3628	Sequence 3628, Ap
55	6	2.3	65	9 US-09-925-297-742	Sequence 742, App
56	6	2.3	65	10 US-09-764-847-647	Sequence 647, App
57	6	2.3	65	12 US-10-029-386-31245	Sequence 31245, A
58	6	2.3	65	15 US-10-092-124-647	Sequence 647, App
59	6	2.3	72	9 US-09-848-288-6	Sequence 6, Appl1
60	6	2.3	73	15 US-10-133-797-75	Sequence 75, Appl1
61	6	2.3	76	9 US-09-822-687-8	Sequence 8, Appl1
62	6	2.3	76	12 US-10-145-586-84	Sequence 84, Appl1
63	6	2.3	77	12 US-10-029-386-29661	Sequence 29661, A
64	6	2.3	78	9 US-09-753-064-3	Sequence 3, Appl1
65	6	2.3	78	9 US-09-753-064-4	Sequence 4, Appl1
66	6	2.3	78	9 US-09-753-064-5	Sequence 5, Appl1
67	6	2.3	78	9 US-09-761-120-13	Sequence 13, Appl1
68	6	2.3	78	9 US-09-761-120-14	Sequence 14, Appl1
69	6	2.3	78	9 US-09-761-120-15	Sequence 15, Appl1
70	6	2.3	78	9 US-09-761-120-16	Sequence 16, Appl1
71	6	2.3	78	9 US-09-761-120-17	Sequence 17, Appl1
72	6	2.3	78	9 US-09-761-120-18	Sequence 18, Appl1
73	6	2.3	78	9 US-09-761-120-19	Sequence 19, Appl1
74	6	2.3	78	9 US-09-761-120-20	Sequence 20, Appl1
75	6	2.3	78	9 US-09-761-120-21	Sequence 21, Appl1
76	6	2.3	78	9 US-09-761-120-22	Sequence 22, Appl1
77	6	2.3	78	9 US-09-761-120-23	Sequence 23, Appl1
78	6	2.3	78	10 US-09-335-325-13	Sequence 13, Appl1
79	6	2.3	78	10 US-09-335-325-14	Sequence 14, Appl1
80	6	2.3	78	10 US-09-335-325-15	Sequence 15, Appl1
81	6	2.3	78	10 US-09-335-325-16	Sequence 16, Appl1
82	6	2.3	78	10 US-09-335-325-17	Sequence 17, Appl1
83	6	2.3	78	10 US-09-335-325-18	Sequence 18, Appl1
84	6	2.3	78	10 US-09-335-325-19	Sequence 19, Appl1
85	6	2.3	78	10 US-09-335-325-20	Sequence 20, Appl1
86	6	2.3	78	10 US-09-335-325-21	Sequence 21, Appl1
87	6	2.3	78	10 US-09-335-325-22	Sequence 22, Appl1
88	6	2.3	78	10 US-09-335-325-23	Sequence 23, Appl1

89 6 2.3 78 12 US-10-267-137-9 Sequence 9, Appl  
90 6 2.3 78 12 US-10-267-137-11 Sequence 11, Appl  
91 6 2.3 78 15 US-10-131-241-13 Sequence 13, Appl  
92 6 2.3 78 15 US-10-131-241-14 Sequence 14, Appl  
93 6 2.3 78 15 US-10-131-241-15 Sequence 15, Appl  
94 6 2.3 78 15 US-10-131-241-16 Sequence 16, Appl  
95 6 2.3 78 15 US-10-131-241-17 Sequence 17, Appl  
96 6 2.3 78 15 US-10-131-241-18 Sequence 18, Appl  
97 6 2.3 78 15 US-10-131-241-19 Sequence 19, Appl  
98 6 2.3 78 15 US-10-131-241-20 Sequence 20, Appl  
99 6 2.3 78 15 US-10-131-241-21 Sequence 21, Appl  
100 6 2.3 78 15 US-10-131-241-22 Sequence 22, Appl

## ALIGNMENTS

RESULT 1  
US-09-084-491A-2  
Sequence 2, Application US/09084491A  
Patent No. US20020061576A1  
GENERAL INFORMATION:  
APPLICANT: MOORE, PAUL A.  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: EBNER, REINHARD  
TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR-LIKE PROTEASE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/084,491A  
FILING DATE: 27-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOKES, ANDERS A.  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PP378  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
US-09-084-491A-2

Query Match 100.0%; Score 263; DB 9; Length 263;  
Best Local Similarity 100.0%; Pred. No. 4,6e-243;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAMVQAFIVSNMMLAEAYSGGCFMNDGHLVREDQTSAPGRLCLNMLDAOSGLASAP 60  
DB 1 MLAMVQAFIVSNMMLAEAYSGGCFMNDGHLVREDQTSAPGRLCLNMLDAOSGLASAP 60  
QY 61 VSGAGNHSYCRNPDEDPGRCWCVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
DB 61 VSGAGNHSYCRNPDEDPGRCWCVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
QY 121 EGGPADEVQVFAFAPANALPARSEAAAVQPVIGISQVRNNSKEXDGLGVLGITMVI 180  
DB 121 EGGPADEVQVFAFAPANALPARSEAAAVQPVIGISQVRNNSKEXDGLGVLGITMVI 180

QY 181 IIAIGAGIILIGYSYKRGKDLKEQHDQKVCEREMORITLPLSAFTNPTEIVDEKTVVHT 240  
DB 181 IIAIGAGIILIGYSYKRGKDLKEQHDQKVCEREMORITLPLSAFTNPTEIVDEKTVVHT 240  
QY 241 SQTVPDPOEGSTPLMGAGTPGA 263  
DB 241 SQTVPDPOEGSTPLMGAGTPGA 263

RESULT 2  
US-10-102-704-2  
Sequence 2, Application US/10102704  
Publication No. US20020164768A1  
GENERAL INFORMATION:  
APPLICANT: Moore et al.  
TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protein  
FILE REFERENCE: PF378C1  
CURRENT APPLICATION NUMBER: US/10/102,704  
PRIOR FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: 09/084,491  
PRIOR FILING DATE: 1998-05-27  
PRIOR APPLICATION NUMBER: 60/048,000  
PRIOR FILING DATE: 1997-05-28  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 263  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-102-704-2

Query Match 100.0%; Score 263; DB 14; Length 263;  
Best Local Similarity 100.0%; Pred. No. 4,6e-243;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAMVQAFIVSNMMLAEAYSGGCFMNDGHLVREDQTSAPGRLCLNMLDAOSGLASAP 60  
DB 1 MLAMVQAFIVSNMMLAEAYSGGCFMNDGHLVREDQTSAPGRLCLNMLDAOSGLASAP 60  
QY 61 VSGAGNHSYCRNPDEDPGRCWCVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
DB 61 VSGAGNHSYCRNPDEDPGRCWCVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
QY 121 EGGPADEVQVFAFAPANALPARSEAAAVQPVIGISQVRNNSKEXDGLGVLGITMVI 180  
DB 121 EGGPADEVQVFAFAPANALPARSEAAAVQPVIGISQVRNNSKEXDGLGVLGITMVI 180  
QY 181 IIAIGAGIILIGYSYKRGKDLKEQHDQKVCEREMORITLPLSAFTNPTEIVDEKTVVHT 240  
DB 181 IIAIGAGIILIGYSYKRGKDLKEQHDQKVCEREMORITLPLSAFTNPTEIVDEKTVVHT 240  
QY 241 SQTVPDPOEGSTPLMGAGTPGA 263  
DB 241 SQTVPDPOEGSTPLMGAGTPGA 263

RESULT 3  
US-10-057-951-2  
Sequence 2, Application US/10057951  
Publication No. US20020177213A1  
GENERAL INFORMATION:  
APPLICANT: Moore et al.  
TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease  
FILE REFERENCE: PF378P1  
CURRENT APPLICATION NUMBER: US/10/057,951  
PRIOR FILING DATE: 2002-01-29  
PRIOR APPLICATION NUMBER: US 09/411,977  
PRIOR FILING DATE: 1999-10-04  
PRIOR APPLICATION NUMBER: US 09/084,491  
PRIOR FILING DATE: 1998-05-27  
PRIOR APPLICATION NUMBER: US 60/048,000  
PRIOR FILING DATE: 1997-05-28

NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Patentin Ver. 3.1  
SEQ ID NO 2  
LENGTH: 263  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-057-951-2

Query Match 100.0%; Score 263; DB 14; Length 263;  
Best Local Similarity 100.0%; Pred. No. 4.6e-243;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLIAVQAFVLSNMLLAEAYGSGGCFWMDNGHLYREDQTSAPGLRCLNMLDAQSGIASAP 60  
DB 1 MLIAVQAFVLSNMLLAEAYGSGGCFWMDNGHLYREDQTSAPGLRCLNMLDAQSGIASAP 60  
QY 61 VSGAGNHSYCRNDEDEPRGWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
DB 61 VSGAGNHSYCRNDEDEPRGWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
QY 121 EGGGADEVQYFAPANALPARSEAAAVQPIGISOVRMNSKEKKDLGTLGYVLGTMVY 180  
DB 121 EGGGADEVQYFAPANALPARSEAAAVQPIGISOVRMNSKEKKDLGTLGYVLGTMVY 180  
QY 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTYVHT 240  
DB 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTYVHT 240  
QY 241 SQTVPDPOEGSTPLMGQAGTPGA 263  
DB 241 SQTVPDPOEGSTPLMGQAGTPGA 263

## RESULT 4

US-10-210-951-44  
Sequence 44, Application US/10210951  
Publication No. US20030170228A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Marsters, Scot A.  
APPLICANT: Pan, James  
APPLICANT: Pitti, Robert M.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Smith, Victoria  
APPLICANT: Stone, Donna M.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR  
FILE REFERENCE: P2931R1C1  
CURRENT APPLICATION NUMBER: US/10/210.951  
PRIOR FILING DATE: 2002-08-02  
PRIOR APPLICATION NUMBER: 60/014699  
PRIOR FILING DATE: 1996-04-01  
PRIOR APPLICATION NUMBER: 60/026943  
PRIOR FILING DATE: 1996-09-23  
PRIOR APPLICATION NUMBER: 60/059121  
PRIOR FILING DATE: 1997-07-17  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/06037  
PRIOR FILING DATE: 1997-10-10  
PRIOR APPLICATION NUMBER: 60/063755  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063045  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063046  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/066511  
PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/066772  
PRIOR FILING DATE: 1997-11-24  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 258  
SEQ ID NO 44  
LENGTH: 263  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-210-951-44

Query Match 95.1%; Score 250; DB 12; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1.2e-230;  
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLIAVQAFVLSNMLLAEAYGSGGCFWMDNGHLYREDQTSAPGLRCLNMLDAQSGIASAP 60  
DB 1 MLIAVQAFVLSNMLLAEAYGSGGCFWMDNGHLYREDQTSAPGLRCLNMLDAQSGIASAP 60  
QY 61 VSGAGNHSYCRNDEDEPRGWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
DB 61 VSGAGNHSYCRNDEDEPRGWCYVSGEAGVPEKRPCEDLRCPETTSQALPAFTTEIOEAS 120  
QY 121 EGGGADEVQYFAPANALPARSEAAAVQPIGISOVRMNSKEKKDLGTLGYVLGTMVY 180  
DB 121 EGGGADEVQYFAPANALPARSEAAAVQPIGISOVRMNSKEKKDLGTLGYVLGTMVY 180  
QY 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTYVHT 240  
DB 181 IIAIGAGIILGYSYKRGKDLKEQHDQKVCEREMQRTLLPLSAFTNPTCEIVDEKTYVHT 240  
QY 241 SQTVPDPOEG 250  
DB 241 SQTVPDPOEG 250

## RESULT 5

US-10-211-884-44  
Sequence 44, Application US/10211884  
Publication No. US20030175900A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Marsters, Scot A.  
APPLICANT: Pan, James  
APPLICANT: Pitti, Robert M.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Smith, Victoria  
APPLICANT: Stone, Donna M.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR  
FILE REFERENCE: P2931R1C1  
CURRENT APPLICATION NUMBER: US/10/211.884  
PRIOR FILING DATE: 2002-08-02  
PRIOR APPLICATION NUMBER: 60/014699  
PRIOR FILING DATE: 1996-04-01  
PRIOR APPLICATION NUMBER: 60/026943  
PRIOR FILING DATE: 1996-09-23  
PRIOR APPLICATION NUMBER: 60/059121  
PRIOR FILING DATE: 1997-07-17  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/062037  
PRIOR FILING DATE: 1997-10-10  
PRIOR APPLICATION NUMBER: 60/063755  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063045  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063046  
PRIOR FILING DATE: 1997-10-24



PRIOR APPLICATION NUMBER: 60/066511  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066772  
PRIOR FILING DATE: 1997-11-24  
Remaining Prior Application data removed - See file Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 258  
SEQ ID NO 44  
LENGTH: 263  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-211-884-44

Query Match 95.1%; Score 250; DB 12; Length 263;  
Best Local Similarity 100.0%; Pred. No. 1,2e-230;  
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLAMVQAFIVSNMLIAEYVSGGCFMNGHLYREDQTSAPAGLRCLNMLDAQGLASAP 60  
DB 1 MLAMVQAFIVSNMLIAEYVSGGCFMNGHLYREDQTSAPAGLRCLNMLDAQGLASAP 60  
QY 61 VSGAGNHSYCNRPDEPRGWCYVSGEAGVPEKRCEDLRCPETTSQALPAFTTEIOEAS 120  
DB 61 VSGAGNHSYCNRPDEPRGWCYVSGEAGVPEKRCEDLRCPETTSQALPAFTTEIOEAS 120  
QY 121 EGGGADDEVQVAPANALPARSEAAAVQVIGISQVRNNSKEKDGLGTVLGTMMVI 180  
DB 121 EGGGADDEVQVAPANALPARSEAAAVQVIGISQVRNNSKEKDGLGTVLGTMMVI 180  
QY 181 IIAAGAGIILGYSYKRGDLEQHDQKVCEREMQRTPLPAFTNPTCEIYDEKTVVHT 240  
DB 181 IIAAGAGIILGYSYKRGDLEQHDQKVCEREMQRTPLPAFTNPTCEIYDEKTVVHT 240  
QY 241 SQTVPDPOEG 250  
DB 241 SQTVPDPOEG 250

## RESULT 6

US-09-864-761-38457  
Sequence 38457, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Femu, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 38457  
LENGTH: 66  
TYPE: PRT  
ORGANISM: Homo sapiens

FEATURE: MAP TO AC002073.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2  
OTHER INFORMATION: SWISSPROT HIT: P39881, EVALU3 3.40e+00  
OTHER INFORMATION: EST\_HUMAN HIT: BE299406.1, EVALU3 5.00e-30  
US-09-864-761-38457

Query Match 25.1%; Score 66; DB 9; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.5e-55;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TTSQLPAFTTEIOEASBGADDEVQVAPANALPARSEAAAVQVIGISQVRNNSKEK 163  
DB 1 TTSQLPAFTTEIOEASBGADDEVQVAPANALPARSEAAAVQVIGISQVRNNSKEK 163  
QY 164 KDGLTL 169  
DB 61 KDGLTL 66

## RESULT 7

US-10-156-761-10667  
Sequence 10667, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 10667  
LENGTH: 457  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-10667

Query Match 3.0%; Score 8; DB 15; Length 457;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 VSGEAGVP 91  
|||||  
DB 352 VSGEAGVP 359

RESULT 8  
US-09-981-151A-55  
; Sequence 55, Application US/09981151A  
; Publication No. US20030212256A1  
; GENERAL INFORMATION:  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Macdougall, John R  
; APPLICANT: Malvankar, Murel M  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Reymann, John A  
; APPLICANT: Stone, David J  
; APPLICANT: Gunther, Erik  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Patuturajan, Meera  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Zehnusen, Bryan D  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spyrek, Kimberly A  
; APPLICANT: Gangoli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Gorman, Linda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-168  
; CURRENT APPLICATION NUMBER: US/09/981,151A  
; CURRENT FILING DATE: 2001-10-16  
; PRIOR APPLICATION NUMBER: 60/241,040  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,058  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,063  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,243  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/242,152  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/242,482  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/242,611  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/242,612  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/242,880  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: 60/242,881  
; PRIOR FILING DATE: 2000-10-24  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 160  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 55  
; LENGTH: 716  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-981-151A-55  
Query Match 3.0%; Score 8; DB 15; Length 716;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 DRRGPMCY 83  
|||||  
DB 163 DRRGPMCY 170

RESULT 9  
US-09-864-761-45155  
; Sequence 45155, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 45155  
; LENGTH: 43  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC020610.6  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.75  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.74  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.82  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.96  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.85  
US-09-864-761-45155  
Query Match 2.7%; Score 7; DB 9; Length 43;

Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 TMMVIT 182  
|||  
Db 26 TMMVIT 32

RESULT 10  
US-10-029-386-27831  
; Sequence 27831, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 27831  
; LENGTH: 69  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO CHR21:66.0  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.83  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
; OTHER INFORMATION: SWISSPROT HIT: P57075, EVALUATE 1.00e-36  
US-10-029-386-27831

Query Match 2.7%; Score 7; DB 12; Length 69;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 VIGISQR 155  
|||  
Db 39 VIGISQR 45

RESULT 11  
US-10-279-579A-18  
; Sequence 18, Application US/10279579A  
; Publication No. US20030177532A1  
; GENERAL INFORMATION:  
; APPLICANT: Burrell, Michael  
; APPLICANT: Coates, Andrew  
; TITLE OF INVENTION: Modification of starch granule size and number  
; FILE REFERENCE: 9341-029  
; CURRENT APPLICATION NUMBER: US/10/279,579A  
; CURRENT FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: 60/346,905  
; PRIOR FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: UK 0125493.7  
; PRIOR FILING DATE: 2001-10-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 18  
; LENGTH: 111  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-279-579A-18

Query Match 2.7%; Score 7; DB 12; Length 111;  
Best Local Similarity 100.0%; Pred. No. 95;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 AAAPV 149  
|||  
Db 94 AAAPV 100

RESULT 12  
US-09-738-626-3788  
; Sequence 3788, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIMAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3788  
; LENGTH: 154  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3788

Query Match 2.7%; Score 7; DB 10; Length 154;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 GIIIGYS 193  
|||  
Db 50 GIIIGYS 56

RESULT 13  
US-09-764-868-727  
; Sequence 727, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT732  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 727  
; LENGTH: 237  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (67)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (119)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-868-727

Query Match 2.7%; Score 7; DB 10; Length 237;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 QPVIGIS 155

Db 215 VIGISQR 221

RESULT 14

US-09-815-242-5680

; Sequence 5680, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ. ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5680

; LENGTH: 258

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-5680

Query Match 2.7%; Score 7; DB 9; Length 256;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 QPVIGIS 153

Db 56 QPVIGIS 62

RESULT 15

US-09-815-242-12173

; Sequence 12173, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ. ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12173

; LENGTH: 300

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12173

Query Match 2.7%; Score 7; DB 9; Length 300;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 QPVIGIS 153

Db 73 QPVIGIS 79

RESULT 16

US-09-943-002-12

; Sequence 12, Application US/09943002

; Patent No. US20020045734A1

; GENERAL INFORMATION:

; APPLICANT: Duncan, Roy

; TITLE OF INVENTION: NOVEL REOVIRUS-DERIVED PROTEINS AND USES THEREFOR

; FILE REFERENCE: 78973-1C

; CURRENT APPLICATION NUMBER: US/09/943,002

; CURRENT FILING DATE: 2001-08-31

; NUMBER OF SEQ. ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 12

; LENGTH: 323

; TYPE: PRT

; ORGANISM: Nelson Bay virus

US-09-943-002-12

Query Match 2.7%; Score 7; DB 9; Length 323;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 AQSGLAS 58

Db 118 AQSGLAS 124

RESULT 17

US-09-815-242-11065

; Sequence 11065, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

```

; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11065
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11065
```

```
Query Match      2.7%; Score 7; DB 9; Length 335;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      182  IAIAGGI 188
          |||||
Db       198  IAIAGGI 204
```

```
RESULT 18
US-09-764-868-850
; Sequence 850, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 850
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-850
```

```
Query Match      2.7%; Score 7; DB 10; Length 399;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      143  AAAYQPV 149
          |||||
Db       234  AAAYQPV 240
```

```
RESULT 19
US-09-815-242-13961
; Sequence 13961, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Maseibeck, Robert
```

```

; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13961
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(465)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13961
```

```
Query Match      2.7%; Score 7; DB 9; Length 465;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      167  GTLGIVL 173
          |||||
Db       190  GTLGIVL 196
```

```
RESULT 20
US-10-156-761-11459
; Sequence 11459, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 11459
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11459
```

Query Match 2.7%; Score 7; DB 15; Length 545;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 APANALP 138  
Db 90 APANALP 96

## RESULT 21

US-09-738-626-5317  
; Sequence 5317, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 5317  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5317

Query Match 2.7%; Score 7; DB 10; Length 562;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 LGTLGYV 172  
Db 511 LGTLGYV 517

## RESULT 22

US-09-946-678-2  
; Sequence 2, Application US/09946678  
; Patent No. US20020106782A1  
; GENERAL INFORMATION:  
; APPLICANT: ITO, Kotaro  
; APPLICANT: UITSUKI, Genryou  
; APPLICANT: KOYAMA, Yasuji  
; TITLE OF INVENTION: Protein Having Glutaminase Activity and Gene Encoding the Same  
; FILE REFERENCE: 0283-0158P  
; CURRENT APPLICATION NUMBER: US/09/946,678  
; CURRENT FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: JP 2000-270371  
; PRIOR FILING DATE: 2000-09-06  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 684  
; TYPE: PRT  
; ORGANISM: Cryptococcus nodansensis  
US-09-946-678-2

Query Match 2.7%; Score 7; DB 10; Length 684;  
Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 DAOSGLA 57  
Db 411 DAOSGLA 417

## RESULT 23

US-09-815-242-13411  
; Sequence 13411, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13411  
; LENGTH: 701  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13411

Query Match 2.7%; Score 7; DB 9; Length 701;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 EIQASE 121  
Db 101 EIQASE 107

## RESULT 24

US-09-738-626-6019  
; Sequence 6019, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
US-09-738-626-6019

APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: Patentin ver. 3.0  
SEQ ID NO 6019  
LENGTH: 706  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6019

Query Match 2.7%; Score 7; DB 10; Length 706;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 ANMLPAR 140  
Db 608 ANMLPAR 614

RESULT 25  
US-09-974-298-129  
Sequence 129, Application US/09974298  
Patent No. US20020156263A1  
GENERAL INFORMATION:  
APPLICANT: Chen, Huel-Mei  
TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER  
FILE REFERENCE: PA-0037 P  
CURRENT APPLICATION NUMBER: US/09/974,298  
CURRENT FILING DATE: 2001-10-04  
PRIOR APPLICATION NUMBER: 60/238,331  
PRIOR FILING DATE: 2000-05-10  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: PBR1 Program  
SEQ ID NO 129  
LENGTH: 937  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURES:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20020156263A1 1331526CD1  
US-09-974-298-129

Query Match 2.7%; Score 7; DB 10; Length 937;  
Best Local Similarity 100.0%; Pred. No. 6.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 HSYCRNP 73  
Db 359 HSYCRNP 365

RESULT 26  
US-10-099-285-72  
Sequence 72, Application US/10099285  
Publication No. US20030105319A1  
GENERAL INFORMATION:  
APPLICANT: Schnepf, H. Ernest  
Wicker, Carol  
Narva, Kenneth E.  
Walz, Michelle  
Stockhoff, Brian  
Muller-Cohn, Judy  
TITLE OF INVENTION: Toxins Active Against Pests  
NUMBER OF SEQUENCES: 105

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: USA  
ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/099,285  
FILING DATE: 15-Mar-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/002,285  
FILING DATE: <Unknown>

APPLICATION NUMBER: US 08/886,615  
FILING DATE: 1-JUL-1997  
APPLICATION NUMBER: US 08/674,002  
FILING DATE: 1-JUL-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M.  
REGISTRATION NUMBER: 39,355

REFERENCE/DOCKET NUMBER: MA-701C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (352) 375-8100  
TELEFAX: (352) 372-5800

INFORMATION FOR SEQ ID NO: 72:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1156 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 72:  
US-10-099-285-72

Query Match 2.7%; Score 7; DB 15; Length 1156;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ASAPVSG 63  
Db 523 ASAPVSG 529

RESULT 27  
US-09-815-242-5064  
Sequence 5064, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyckind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26

```
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; LENGTH: 2472
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5064
```

```
Query Match      2.7%: Score 7; DB 9; Length 2472;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      50 LDAQSGL 56
        |||||
Db       1753 LDAQSGL 1759
```

```
RESULT 28
US-09-792-286-183
; Sequence 183, Application US/09792286
; Publication No. US20030092609A1
; GENERAL INFORMATION:
; APPLICANT: LARSEN, BJARNE D.
; APPLICANT: PETERSEN, JORGEN S.
; APPLICANT: MEIER, EDDI
; APPLICANT: KJOLBYE, ANNE L.
; APPLICANT: JORGENSEN, NIKLAS R.
; APPLICANT: NIELSEN, MORTEN S.
; APPLICANT: MARTINS, JAMES B.
; APPLICANT: HOLSTEIN-RATHLOU, NEILS-HENRIK
; TITLE OF INVENTION: NOVEL ANTIRHYTHMIC PEPTIDES
; FILE REFERENCE: 55461-C (45487)
; CURRENT APPLICATION NUMBER: US/09/792,286
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 299
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 183
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-792-286-183
```

```
Query Match      2.3%: Score 6; DB 11; Length 7;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      61 VSGAGN 66
        |||||
Db       2 VSGAGN 7
```

```
RESULT 29
US-09-792-286-258
; Sequence 258, Application US/09792286
; Publication No. US20030092609A1
; GENERAL INFORMATION:
; APPLICANT: LARSEN, BJARNE D.
; APPLICANT: PETERSEN, JORGEN S.
; APPLICANT: MEIER, EDDI
; APPLICANT: KJOLBYE, ANNE L.
; APPLICANT: JORGENSEN, NIKLAS R.
; APPLICANT: NIELSEN, MORTEN S.
```

```
; APPLICANT: MARTINS, JAMES B.
; APPLICANT: HOLSTEIN-RATHLOU, NEILS-HENRIK
; TITLE OF INVENTION: NOVEL ANTIRHYTHMIC PEPTIDES
; FILE REFERENCE: 55461-C (45487)
; CURRENT APPLICATION NUMBER: US/09/792,286
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 299
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 258
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-792-286-258
```

```
Query Match      2.3%: Score 6; DB 11; Length 7;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      61 VSGAGN 66
        |||||
Db       2 VSGAGN 7
```

```
RESULT 30
US-10-206-699-74
; Sequence 74, Application US/10206699
; Publication No. US20030100510A1
; GENERAL INFORMATION:
; APPLICANT: Sundaramoorthy, M.
; APPLICANT: Hudson, B.
; TITLE OF INVENTION: Crystallized structure of Type IV Collagen NC1 Domain Hexamer
; FILE REFERENCE: MBHB 01-1017
; CURRENT APPLICATION NUMBER: US/10/206,699
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/308,523
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/351,289
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/366,854
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/385,362
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-206-699-74
```

```
Query Match      2.3%: Score 6; DB 15; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      165 DGLTIG 170
        |||||
Db       11 DGLTIG 16
```

```
RESULT 31
US-10-032-221B-28
; Sequence 28, Application US/10032221B
; Publication No. US2003014481A1
; GENERAL INFORMATION:
; APPLICANT: Kaliluri, Raghubram
; TITLE OF INVENTION: ANTI-ANGIOGENIC PROTEINS AND FRAGMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 2312/2082B (formerly 1440.1027-016)
; CURRENT APPLICATION NUMBER: US/10/032,221B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/US01/00565
```



```

; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/625,191
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/543,371
; PRIOR FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: US 09/479,118
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/335,224
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/126,175
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/089,689
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: T2 (amino acids 53-72 of SEQ ID NO:10)
US-10-032-221B-28
```

```

Query Match          2.3%; Score 6; DB 12; Length 20;
Best Local Similarity 100.0%; Pred.No.1.9e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;
```

```
Qy      165 DLGTLG 170
Db      8 DLGTLG 13
```

## RESULT 32

```

US-09-864-761-35595
; Sequence 35595, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35595
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009474.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.5
; OTHER INFORMATION: EST_HUMAN HIT: BF035327.1, EVALU 7.00e-03
US-09-864-761-35595
```

```

Query Match          2.3%; Score 6; DB 9; Length 27;
Best Local Similarity 100.0%; Pred.No.2.4e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;
```

```
Qy      219 PLSAFT 224
Db      18 PLSAFT 23
```

## RESULT 33

```

US-09-864-761-48415
; Sequence 48415, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48415
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL096704.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
; OTHER INFORMATION: EST_HUMAN HIT: AA250859.1, EVALU8 7.80e-02
US-09-864-761-48415
```

```

Query Match      2.3%; Score 6; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      159 NSXEKK 164
      |||||
Db      19 NSXEKK 24

RESULT 34
US-10-160-162-121
; Sequence 121, Application US/10160162
; Publication No. US20030166541A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: e3 Human Secreted Proteins
; FILE REFERENCE: P2012P2
; CURRENT APPLICATION NUMBER: US/10/160,162
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: 60/295,558
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/236,557
; PRIOR FILING DATE: 1999-01-26
; PRIOR APPLICATION NUMBER: PCT/US98/15949
; PRIOR FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,212
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,209
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,234
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,218
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,214
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,236
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,215
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,211
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,217
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,213
```

```

; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/055,968
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,969
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,972
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/056,561
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,534
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,729
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,543
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,727
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,554
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,730
; PRIOR FILING DATE: 1997-08-19
; NUMBER OF SEQ ID NOS: 353
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals stop translation
US-10-160-162-121
```

```

Query Match      2.3%; Score 6; DB 12; Length 34;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      56 LASAPV 61
      |||||
Db      21 LASAPV 26

RESULT 35
US-09-820-649-121
; Sequence 121, Application US/09820649
; Publication No. US20030199683A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: e3 Human Secreted Proteins
; FILE REFERENCE: P2012P1
; CURRENT APPLICATION NUMBER: US/09/820,649
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US/09/236,557
; PRIOR FILING DATE: 1999-01-26
; PRIOR APPLICATION NUMBER: PCT/US98/15949
; PRIOR FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,212
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,209
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,234
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,218
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,214
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,236
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,215
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/054,211
; PRIOR FILING DATE: 1997-07-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
```

```
; NUMBER OF SEQ ID NOS: 353
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals stop translation
US-09-820-649-121
```

```
Query Match      2.3%; Score 6; DB 12; Length 34;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      56 LASAPV 61
        |||||
Db       21 LASAPV 26
```

```
RESULT 36
US-10-106-698-8286
; Sequence 8286, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8286
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-8286
```

```
Query Match      2.3%; Score 6; DB 15; Length 37;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      256 GQAGTP 261
        |||||
Db       9 GQAGTP 14
```

```
RESULT 37
US-09-864-761-38182
; Sequence 38182, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Fenn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
```

```
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38182
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035662.23
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EST_HUMAN HIT: AA147464.1, EVALU6 6.00e-19
; OTHER INFORMATION: SWISSPROT HIT: P48809, EVALU6 4.70e+00
US-09-864-761-38182
```

```
Query Match      2.3%; Score 6; DB 9; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      257 QAGTPG 262
        |||||
Db       4 QAGTPG 9
```

```
RESULT 38
```

```
US-09-983-802-589
; Sequence 589, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
```

```
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 589
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-802-589

Query Match
Best Local Similarity 2.3%; Score 6; DB 11; Length 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      80 PWCYVS 85
Db       32 PWCYVS 37

RESULT 39
US-09-864-761-34850
; Sequence 34850, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
```

```

; SEQ ID NO 34850
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109933.5
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 15
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.4
; OTHER INFORMATION: EST HUMAN HIT: AVE694621.1, EVALUE 2.00e-20
; OTHER INFORMATION: SWISSPROT HIT: P00747, EVALUE 1.00e-21
US-09-864-761-34850
```

```

Query Match      2.3%; Score 6; DB 9; Length 42;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      69 YCRNPD 74
      |||||
Db      10 YCRNPD 15
```

```

RESULT 40
US-10-106-698-8514
; Sequence 8514, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8514
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (23)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (28)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (33)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-8514
```

```

Query Match      2.3%; Score 6; DB 15; Length 46;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      165 DLGTLG 170
      |||||
Db      9 DLGTLG 14
```

```

RESULT 41
US-09-864-761-40317
; Sequence 40317, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-09-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40317
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL162458.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EST_HUMAN HIT: BB831812.1, EVALUE 8.80e-02
US-09-864-761-40317
```

Query Match 2.3%; Score 6; DB 9; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 GLASAP 60  
 |||||  
 DB 42 GLASAP 47

RESULT 42  
 US-09-864-761-37671  
 ; Sequence 37671, Application US/09864761  
 ; Patent No. US2002048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 ; FILE REFERENCE: Aemica-X-1  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-06-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 37671  
 ; LENGTH: 51  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AL021579.1  
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4  
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5  
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98  
 ; OTHER INFORMATION: EST\_HUMAN HIT: AUI26907.1, EVALUUE 7.00e-14  
 US-09-864-761-37671

Query Match 2.3%; Score 6; DB 9; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 AOSGLA 57  
 |||||  
 DB 10 AOSGLA 15

RESULT 43  
 US-10-029-386-29128  
 ; Sequence 29128, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AEMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 29128  
 ; LENGTH: 52  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO CHR7.1  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3  
 ; OTHER INFORMATION: SWISSPROT HIT: Q02643, EVALUUE 1.60e+00  
 US-10-029-386-29128

Query Match 2.3%; Score 6; DB 12; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 ALPARS 141  
 |||||  
 DB 27 ALPARS 32

RESULT 44  
 US-10-218-102-371  
 ; Sequence 371, Application US/10218102  
 ; Publication No. US20030130827A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bentzien, Joerg  
 ; APPLICANT: Dabiyat, Bassil I.  
 ; APPLICANT: Desjarlais, John R.  
 ; APPLICANT: Hayes, Robert J.  
 ; APPLICANT: Vielmeyer, Jost  
 ; TITLE OF INVENTION: Protein Design Automation for Protein Libraries  
 ; FILE REFERENCE: A-67229-11/RT/RMS/RMK  
 ; CURRENT APPLICATION NUMBER: US/10/218,102  
 ; CURRENT FILING DATE: 2002-08-12  
 ; PRIOR APPLICATION NUMBER: US 09/927,790  
 ; PRIOR FILING DATE: 2001-08-10

```

; PRIOR APPLICATION NUMBER: US 60/311,545
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US 60/324,899
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/351,937
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 60/352,103
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 432
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 371
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-218-102-371
```

```
Query Match      2.3%; Score 6; DB 16; Length 52;
Best Local Similarity 100.0%; Pred. No. 4,4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      196 RGKDLK 201
Db      18 RGKDLK 23
```

```

RESULT 45
; Sequence 588, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
```

```

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 588
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-802-588
```

```
Query Match      2.3%; Score 6; DB 11; Length 53;
Best Local Similarity 100.0%; Pred. No. 4,4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      80 PWCYVS 85
Db      46 PWCYVS 51
```

```

RESULT 46
; Sequence 31083, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31083
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008901.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.55
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
```

```
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
; OTHER INFORMATION: SWISSPROT HIT: P53578, EVALU8.00e+00
US-10-029-386-31083

Query Match
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      86 GEAGVP 91
      |||||
Db      23 GEAGVP 28

RESULT 47
US-10-007-280A-184
; Sequence 184, Application US/10007280A
; Publication No. US20030059784A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; APPLICANT: Salceda, Susana
; APPLICANT: Changhai, Liu
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0257
; CURRENT APPLICATION NUMBER: US/10/007,280A
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,640
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 184
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-007-280A-184

Query Match
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      57 ASAPVS 62
      |||||
Db      33 ASAPVS 38

RESULT 48
US-09-879-957-215
; Sequence 215, Application US/0987957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US20020034755A1h
; KAY, Brian K.
; FOWKES, Dana M.
; McCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
```

```
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 215:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 59 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 215:
US-09-879-957-215

Query Match
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      231 VDEKTV 236
      |||||
Db      29 VDEKTV 34

RESULT 49
US-09-864-761-47913
; Sequence 47913, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
```



```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47913
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008073.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
; OTHER INFORMATION: EST HUMAN HIT: AW206547.1, EVALUATE 3.00e-30
; OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 2.00e-09
US-09-864-761-47913
```

```
Query Match      2.3%; Score 6; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      231 VDEKTV 236
Db      1 VDEKTV 6
```

```
RESULT 50
US-10-156-761-7793
; Sequence 7793, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7793
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7793
```

```
Query Match      2.3%; Score 6; DB 15; Length 60;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      184 IGAGII 189
Db      46 IGAGII 51
```

```
RESULT 51
US-09-912-787-77
; Sequence 77, Application US/09912787
; Patent No. US20020119149A1
; GENERAL INFORMATION:
```

```
; APPLICANT: JAKOBSEN, Bent Karsten
; APPLICANT: BOULTER, Jonathan Michael
; TITLE OF INVENTION: Multivalent T Cell Receptor Complexes
; FILE REFERENCE: 102286.410DIV
; CURRENT APPLICATION NUMBER: US/09/912,787
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: US 09/334,969
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: PCT/GB99 01583
; PRIOR FILING DATE: 1999-05-19
; PRIOR APPLICATION NUMBER: GB 9810759.2
; PRIOR FILING DATE: 1998-05-19
; PRIOR APPLICATION NUMBER: GB 9821129.5
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 77
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence of
; OTHER INFORMATION: the c-Fos/BiRa biotinylation tag fusion partner
; OTHER INFORMATION: used for TCR beta chains.
US-09-912-787-77
```

```
Query Match      2.3%; Score 6; DB 10; Length 61;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      19 AYSGSG 24
Db      42 AYSGSG 47
```

```
RESULT 52
US-10-014-326-63
; Sequence 63, Application US/10014326
; Publication No. US20020142389A1
; GENERAL INFORMATION:
; APPLICANT: JAKOBSEN, Bent Karsten
; APPLICANT: BELL, John Irving
; APPLICANT: GAO, George Fu
; APPLICANT: WILCOX, Benjamin Ernest
; APPLICANT: BOULTER, Jonathan Michael
; TITLE OF INVENTION: Soluble T Cell Receptor
; FILE REFERENCE: 102286.409
; CURRENT APPLICATION NUMBER: US/10/014,326
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/335,087
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: GB/9810759.2
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-05-19
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: amino acid
; OTHER INFORMATION: sequence of the c-Fos-BiRa biotinylation tag
; OTHER INFORMATION: fusion partner used for TCR beta chains.
US-10-014-326-63
```

```
Query Match      2.3%; Score 6; DB 14; Length 61;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      19 AYSGSG 24
Db      42 AYSGSG 47
```

```
RESULT 53
US-09-864-761-40363
; Sequence 40363, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40363
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL158176.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EST_HUMAN HIT: BE897259.1, EVALU8 8.00e-08
US-09-864-761-40363

Query Match      2.3%: Score 6; DB 9; Length 62;
Best Local Similarity 100.0%; Pred.No.5.1e+02;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
```

```
QY      38 TSPAPG 43
Db      2 TSPAPG 7

RESULT 54
US-09-738-626-3628
; Sequence 3628, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 248-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3628
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3628

Query Match      2.3%: Score 6; DB 10; Length 63;
Best Local Similarity 100.0%; Pred.No.5.2e+02;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      54 SGLASA 59
Db      39 SGLASA 44

RESULT 55
US-09-925-297-742
; Sequence 742, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 742
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-742

Query Match      2.3%: Score 6; DB 9; Length 65;
Best Local Similarity 100.0%; Pred.No.5.3e+02;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
```

QY 187 GILIGY 192  
Db 4 GILIGY 9

RESULT 56  
US-09-764-847-647  
; Sequence 647, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764, 847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 647  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-847-647

Query Match 2.3%; Score 6; DB 10; Length 65;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 GILIGY 192  
Db 4 GILIGY 9

RESULT 57  
US-10-029-386-31245  
; Sequence 31245, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; FILE REFERENCE: ABOICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 31245  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC016605.4  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4  
; OTHER INFORMATION: SWISSPROT HIT: O51532, EVALU8 7.40e+00  
US-10-029-386-31245

Query Match 2.3%; Score 6; DB 12; Length 65;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 LGYVIG 174  
Db 43 LGYVIG 48

RESULT 58

US-10-092-154-647  
; Sequence 647, Application US/10092154  
; Publication No. US20030054375A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009C1  
; CURRENT APPLICATION NUMBER: US/10/092,154  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2003  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 647  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-092-154-647

Query Match 2.3%; Score 6; DB 15; Length 65;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 GILIGY 192  
Db 4 GILIGY 9

RESULT 59  
US-09-848-288-6  
; Sequence 6, Application US/09848288  
; Patent No. US20020051984A1  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Kingle Domain-Containing Polynucleotides, Polypeptides, and P  
; FILE REFERENCE: PTO22P1  
; CURRENT APPLICATION NUMBER: US/09/848,288  
; CURRENT FILING DATE: 2001-05-04  
; PRIOR APPLICATION NUMBER: PCT/US00/30664  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/164,853  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-848-288-6

Query Match 2.3%; Score 6; DB 9; Length 72;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
Db 41 YCRNPD 46

RESULT 60  
US-10-133-797-75  
; Sequence 75, Application US/10133797  
; Publication No. US20030109021A1  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Shujian  
; APPLICANT: Chen, Jian  
; APPLICANT: Feder, John  
; APPLICANT: Lee, Liana  
; APPLICANT: Krystek, Stanley  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY  
; FILE REFERENCE: D0141NP  
; CURRENT APPLICATION NUMBER: US/10/133,797  
; CURRENT FILING DATE: 2002-04-26

```

; PRIOR APPLICATION NUMBER: US 60/286,764
; PRIOR FILING DATE: 2001-04-26
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 75
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-797-75

```

```

Query Match      2.3%; Score 6; DB 15; Length 73;
Best Local Similarity 100.0%; Pred.No. 5.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      134 ANALPA 139
      |||||
Db      48 ANALPA 53

```

```

RESULT 61
US-09-822-687-8
; Sequence 8, Application US/09822687
; Patent No. US2002076753A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria Alexandra
; TITLE OF INVENTION: 31939, A NOVEL HUMAN LEUCINE-RICH REPEAT
; FILE REFERENCE: 10448-038001
; CURRENT APPLICATION NUMBER: US/09/822,687
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/193,919
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-822-687-8

```

```

Query Match      2.3%; Score 6; DB 9; Length 76;
Best Local Similarity 100.0%; Pred.No. 6.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      190 LGSYK 195
      |||||
Db      34 LGSYK 39

```

```

RESULT 62
US-10-145-586-84
; Sequence 84, Application US/10145586
; Publication No. US20030138890A1
; GENERAL INFORMATION:
; APPLICANT: Alexandra Glucksmann, Maria
; APPLICANT: Silos-Santiago, Imaculada
; APPLICANT: M. Galvin, Katherine
; APPLICANT: Welch, Nadine
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Kapeller-Liebermann, Rosana
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTOR FAMILY MEMBERS,
; TITLE OF INVENTION: HUMAN THIOREDOXIN FAMILY MEMBERS, HUMAN LEUCINE-RICH
; TITLE OF INVENTION: REPEAT FAMILY MEMBERS, AND HUMAN RINGFINGER FAMILY MEMBER
; FILE REFERENCE: 10448-188001
; CURRENT APPLICATION NUMBER: US/10/145,586
; CURRENT FILING DATE: 2002-05-14
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84

```

```

; LENGTH: 76
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-10-145-586-84

```

```

Query Match      2.3%; Score 6; DB 12; Length 76;
Best Local Similarity 100.0%; Pred.No. 6.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      190 LGSYK 195
      |||||
Db      34 LGSYK 39

```

```

RESULT 63
US-10-029-386-29661
; Sequence 29661, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29661
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR19.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.48
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.46
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.46
; OTHER INFORMATION: SWISSPROT HIT: O35984, EVALUE 3.00e-17
US-10-029-386-29661

```

```

Query Match      2.3%; Score 6; DB 12; Length 77;
Best Local Similarity 100.0%; Pred.No. 6.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      13 NMLAE 18
      |||||
Db      8 NMLAE 13

```

```

RESULT 64
US-09-753-064-3
; Sequence 3, Application US/09753064
; Patent No. US2001001664A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yihai
; TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Jones & Askew, LLP
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/753,064  
FILING DATE: 29-Dec-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/763,528  
FILING DATE: 12-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05940-0251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..78  
OTHER INFORMATION: /note="Kring 2 - Figure 3"  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-753-064-3  
Query Match 2.3%; Score 6; DB 9; Length 78;  
Best Local Similarity 100.0%; Pred. No. 6.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 YCRNPD 74  
Db 49 YCRNPD 54  
RESULT 65  
US-09-753-064-4  
Sequence 4, Application US/09753064  
Patent No. US2001001644A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yihai  
Folkman, M. Judah  
TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/753,064  
FILING DATE: 29-Dec-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/763,528  
FILING DATE: 12-DEC-1996  
ATTORNEY/AGENT INFORMATION:

NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05940-0251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..78  
OTHER INFORMATION: /note="Kring 3 - Figure 3"  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-753-064-4  
Query Match 2.3%; Score 6; DB 9; Length 78;  
Best Local Similarity 100.0%; Pred. No. 6.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 YCRNPD 74  
Db 49 YCRNPD 54  
RESULT 66  
US-09-753-064-5  
Sequence 5, Application US/09753064  
Patent No. US2001001644A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yihai  
Folkman, M. Judah  
TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/753,064  
FILING DATE: 29-Dec-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/763,528  
FILING DATE: 12-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05940-0251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

```
/
/
/ MOLECULE TYPE: linear
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/ FEATURE:
/ NAME/KEY: Protein
/ LOCATION: 1..78
/ OTHER INFORMATION: /note= "Kring1e 4 - Figure 3"
US-09-753-064-5
SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 78;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74
Db 49 YCRNPD 54

RESULT 67
US-09-761-120-13
Sequence 13, Application US/09761120
Patent No. US20020037847A1
GENERAL INFORMATION:
APPLICANT: Folkman, M. Judah
APPLICANT: O'Reilly, Michael
TITLE OF INVENTION: Nucleic Acids Encoding Kring1e 1-5 Region Fragments of Plasminogen
FILE REFERENCE: 05940-0151 (43171-252068)
CURRENT APPLICATION NUMBER: US/09/761,120
CURRENT FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 09/309,821
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 08/866,735
PRIOR FILING DATE: 1997-05-30
NUMBER OF SEQ ID NOS: 47
SOFTWARE: PatentIn version 3.0
SEQ ID NO 13
LENGTH: 78
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Kring1e 2
US-09-761-120-13

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 78;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74
Db 49 YCRNPD 54

RESULT 68
US-09-761-120-14
Sequence 14, Application US/09761120
Patent No. US20020037847A1
GENERAL INFORMATION:
APPLICANT: Folkman, M. Judah
APPLICANT: O'Reilly, Michael
TITLE OF INVENTION: Nucleic Acids Encoding Kring1e 1-5 Region Fragments of Plasminogen
FILE REFERENCE: 05940-0151 (43171-252068)
CURRENT APPLICATION NUMBER: US/09/761,120
CURRENT FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 09/309,821
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 08/866,735
PRIOR FILING DATE: 1997-05-30
NUMBER OF SEQ ID NOS: 47
SOFTWARE: PatentIn version 3.0
```

```
/
/
/ SEQ ID NO 14
/ LENGTH: 78
/ TYPE: PRT
/ ORGANISM: Macaca sp.
/ FEATURE:
/ NAME/KEY: misc.feature
/ OTHER INFORMATION: Kring1e 2
US-09-761-120-14

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 78;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74
Db 49 YCRNPD 54

RESULT 69
US-09-761-120-15
Sequence 15, Application US/09761120
Patent No. US20020037847A1
GENERAL INFORMATION:
APPLICANT: Folkman, M. Judah
APPLICANT: O'Reilly, Michael
TITLE OF INVENTION: Nucleic Acids Encoding Kring1e 1-5 Region Fragments of Plasminogen
FILE REFERENCE: 05940-0151 (43171-252068)
CURRENT APPLICATION NUMBER: US/09/761,120
CURRENT FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 09/309,821
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 08/866,735
PRIOR FILING DATE: 1997-05-30
NUMBER OF SEQ ID NOS: 47
SOFTWARE: PatentIn version 3.0
SEQ ID NO 15
LENGTH: 78
TYPE: PRT
ORGANISM: Sus sp.
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Kring1e 2
US-09-761-120-15

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 78;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74
Db 49 YCRNPD 54

RESULT 70
US-09-761-120-16
Sequence 16, Application US/09761120
Patent No. US20020037847A1
GENERAL INFORMATION:
APPLICANT: Folkman, M. Judah
APPLICANT: O'Reilly, Michael
TITLE OF INVENTION: Nucleic Acids Encoding Kring1e 1-5 Region Fragments of Plasminogen
FILE REFERENCE: 05940-0151 (43171-252068)
CURRENT APPLICATION NUMBER: US/09/761,120
CURRENT FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 09/309,821
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 08/866,735
PRIOR FILING DATE: 1997-05-30
NUMBER OF SEQ ID NOS: 47
SOFTWARE: PatentIn version 3.0
SEQ ID NO 16
LENGTH: 78
TYPE: PRT
```

```

; ORGANISM: Bos sp.
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Kringle 2
US-09-761-120-16
```

```
Query Match          2.3%; Score 6; DB 9; Length 78;
Best Local Similarity 100.0%; Pred. No. 6.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 69 YCRNPD 74
    |||||
Db 49 YCRNPD 54
```

```
RESULT 71
US-09-761-120-17
; Sequence 17, Application US/09761120
; Patent No. US20020037847A1
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Nucleic Acids Encoding Kringle 1-5 Region Fragments of Plasminogen
; FILE REFERENCE: 05940-0151 (43171-252068)
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US/09/761,120
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 09/309,821
; PRIOR FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Kringle 3
US-09-761-120-17
```

```
Query Match          2.3%; Score 6; DB 9; Length 78;
Best Local Similarity 100.0%; Pred. No. 6.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 69 YCRNPD 74
    |||||
Db 49 YCRNPD 54
```

```
RESULT 72
US-09-761-120-18
; Sequence 18, Application US/09761120
; Patent No. US20020037847A1
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Nucleic Acids Encoding Kringle 1-5 Region Fragments of Plasminogen
; FILE REFERENCE: 05940-0151 (43171-252068)
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US/09/761,120
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 09/309,821
; PRIOR FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```

; OTHER INFORMATION: Kringle 3
US-09-761-120-18
```

```
Query Match          2.3%; Score 6; DB 9; Length 78;
Best Local Similarity 100.0%; Pred. No. 6.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 69 YCRNPD 74
    |||||
Db 49 YCRNPD 54
```

```
RESULT 73
US-09-761-120-19
; Sequence 19, Application US/09761120
; Patent No. US20020037847A1
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Nucleic Acids Encoding Kringle 1-5 Region Fragments of Plasminogen
; FILE REFERENCE: 05940-0151 (43171-252068)
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US/09/761,120
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 09/309,821
; PRIOR FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Macaca sp.
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Kringle 3
US-09-761-120-19
```

```
Query Match          2.3%; Score 6; DB 9; Length 78;
Best Local Similarity 100.0%; Pred. No. 6.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 69 YCRNPD 74
    |||||
Db 49 YCRNPD 54
```

```
RESULT 74
US-09-761-120-20
; Sequence 20, Application US/09761120
; Patent No. US20020037847A1
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Nucleic Acids Encoding Kringle 1-5 Region Fragments of Plasminogen
; FILE REFERENCE: 05940-0151 (43171-252068)
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US/09/761,120
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 09/309,821
; PRIOR FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Sus sp.
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Kringle 3
US-09-761-120-20
```

```
Query Match          2.3%; Score 6; DB 9; Length 78;
Best Local Similarity 100.0%; Pred.No. 6.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74
   |||||
Db 49 YCRNPD 54
```

```
RESULT 75
US-09-761-120-21
; Sequence 21, Application US/09761120
; Patent No. US20020037847A1
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Nucleic Acids Encoding Kringle 1-5 Region Fragments of Plasminogen
; FILE REFERENCE: 05940-0151 (43171-252068)
; CURRENT APPLICATION NUMBER: US/09/761,120
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 09/309,821
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 08/866,735
; PRIOR FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Bos sp.
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Kringle 3
US-09-761-120-21
```

```
Query Match          2.3%; Score 6; DB 9; Length 78;
Best Local Similarity 100.0%; Pred.No. 6.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 69 YCRNPD 74
   |||||
Db 49 YCRNPD 54
```

Search completed: November 25, 2003, 13:46:41  
Job time : 33 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:38:54 ; Search time 21 Seconds

(without alignments)  
529.893 Million cell updates/sec

Title: US-10-057-951-2

Perfect score: 263  
Sequence: 1 MLIAWQAFLYSNMILAEAY.....PVDPEGSTPLMGAGTGA 263

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	263	100.0	263	US-09-411-977-2	Sequence 2, Appli
2	8	3.0	716	US-08-766-982-1	Sequence 1, Appli
3	8	3.0	716	US-09-296-219-1	Sequence 846, App
4	7	2.7	182	US-09-198-452A-846	Sequence 5978, Ap
5	7	2.7	204	US-09-107-532A-5978	Sequence 849, App
6	7	2.7	228	US-09-198-452A-849	Sequence 18108, A
7	7	2.7	341	US-09-252-991A-18308	Sequence 5053, Ap
8	7	2.7	353	US-09-107-532A-5053	Sequence 4004, Ap
9	7	2.7	356	US-09-107-532A-4004	Sequence 6936, Ap
10	7	2.7	356	US-09-107-532A-6936	Sequence 29970, A
11	7	2.7	381	US-09-252-991A-29970	Sequence 28999, A
12	7	2.7	413	US-09-252-991A-28999	Sequence 27, Appli
13	7	2.7	497	US-08-009-075-4	Sequence 27, Appli
14	7	2.7	569	US-08-467-822-27	Sequence 27, Appli
15	7	2.7	569	US-08-432-697-27	Sequence 27, Appli
16	7	2.7	569	US-08-466-248-27	Sequence 27, Appli
17	7	2.7	575	US-08-403-866-7	Sequence 7, Appli
18	7	2.7	625	US-08-532-547-7	Sequence 7, Appli
19	7	2.7	625	US-08-532-547-9	Sequence 7, Appli
20	7	2.7	625	US-09-019-809-7	Sequence 7, Appli
21	7	2.7	625	US-09-019-809-9	Sequence 9, Appli
22	7	2.7	625	US-09-471-177-7	Sequence 9, Appli
23	7	2.7	625	US-09-471-177-9	Sequence 9, Appli
24	7	2.7	684	US-09-946-678-2	Sequence 72, Appli
25	7	2.7	800	US-08-469-537A-72	Sequence 78, Appli
26	7	2.7	800	US-08-469-537A-78	Sequence 8, Appli
27	7	2.7	831	US-09-269-861A-8	

28	7	2.7	937	US-08-469-537A-105	Sequence 105, App
29	7	2.7	1156	US-09-002-285-72	Sequence 72, Appli
30	7	2.7	1156	US-09-589-477-72	Sequence 72, Appli
31	7	2.7	1157	US-08-532-547-5	Sequence 5, Appli
32	7	2.7	1157	US-08-379-656B-5	Sequence 5, Appli
33	7	2.7	1157	US-08-445-838-5	Sequence 5, Appli
34	7	2.7	1157	US-09-019-809-5	Sequence 5, Appli
35	7	2.7	1157	US-09-471-177-5	Sequence 5, Appli
36	7	2.7	1247	US-09-252-991A-52960	Sequence 32960, A
37	6	2.3	7	US-09-192-012-8	Sequence 8, Appli
38	6	2.3	8	US-08-267-092A-12	Sequence 12, Appli
39	6	2.3	8	US-08-468-161-12	Sequence 12, Appli
40	6	2.3	8	PCT-US95-08156-12	Sequence 12, Appli
41	6	2.3	38	US-09-348-953-4	Sequence 4, Appli
42	6	2.3	39	US-09-227-357-589	Sequence 4, Appli
43	6	2.3	46	US-09-348-953-6	Sequence 6, Appli
44	6	2.3	53	US-09-227-357-588	Sequence 588, App
45	6	2.3	59	US-08-630-915A-215	Sequence 215, App
46	6	2.3	66	US-09-107-532A-6727	Sequence 6727, App
47	6	2.3	73	US-09-107-532A-4213	Sequence 4213, Ap
48	6	2.3	78	US-08-612-788-13	Sequence 13, Appli
49	6	2.3	78	US-08-612-788-14	Sequence 13, Appli
50	6	2.3	78	US-08-612-788-15	Sequence 15, Appli
51	6	2.3	78	US-08-612-788-16	Sequence 16, Appli
52	6	2.3	78	US-08-612-788-17	Sequence 17, Appli
53	6	2.3	78	US-08-612-788-18	Sequence 18, Appli
54	6	2.3	78	US-08-612-788-19	Sequence 19, Appli
55	6	2.3	78	US-08-612-788-20	Sequence 20, Appli
56	6	2.3	78	US-08-612-788-21	Sequence 21, Appli
57	6	2.3	78	US-08-612-788-22	Sequence 22, Appli
58	6	2.3	78	US-08-612-788-23	Sequence 23, Appli
59	6	2.3	78	US-08-763-528A-3	Sequence 3, Appli
60	6	2.3	78	US-08-763-528A-4	Sequence 4, Appli
61	6	2.3	78	US-08-763-528A-5	Sequence 5, Appli
62	6	2.3	78	US-09-066-028-13	Sequence 13, Appli
63	6	2.3	78	US-09-066-028-14	Sequence 14, Appli
64	6	2.3	78	US-09-066-028-15	Sequence 15, Appli
65	6	2.3	78	US-09-066-028-16	Sequence 16, Appli
66	6	2.3	78	US-09-066-028-17	Sequence 17, Appli
67	6	2.3	78	US-09-066-028-18	Sequence 18, Appli
68	6	2.3	78	US-09-066-028-19	Sequence 19, Appli
69	6	2.3	78	US-09-066-028-20	Sequence 20, Appli
70	6	2.3	78	US-09-066-028-21	Sequence 21, Appli
71	6	2.3	78	US-09-066-028-22	Sequence 22, Appli
72	6	2.3	78	US-09-066-028-23	Sequence 23, Appli
73	6	2.3	78	US-09-335-325-11	Sequence 11, Appli
74	6	2.3	78	US-09-335-325-15	Sequence 15, Appli
75	6	2.3	78	US-09-335-325-16	Sequence 16, Appli
76	6	2.3	78	US-09-335-325-17	Sequence 17, Appli
77	6	2.3	78	US-09-335-325-18	Sequence 18, Appli
78	6	2.3	78	US-09-335-325-19	Sequence 19, Appli
79	6	2.3	78	US-09-335-325-20	Sequence 20, Appli
80	6	2.3	78	US-09-335-325-21	Sequence 21, Appli
81	6	2.3	78	US-09-335-325-22	Sequence 22, Appli
82	6	2.3	78	US-09-335-325-23	Sequence 23, Appli
83	6	2.3	78	US-09-348-953-3	Sequence 3, Appli
84	6	2.3	79	US-08-612-788-7	Sequence 7, Appli
85	6	2.3	79	US-08-612-788-8	Sequence 8, Appli
86	6	2.3	79	US-08-612-788-9	Sequence 9, Appli
87	6	2.3	79	US-08-612-788-10	Sequence 10, Appli
88	6	2.3	79	US-08-612-788-11	Sequence 11, Appli
89	6	2.3	79	US-08-763-528A-1	Sequence 1, Appli
90	6	2.3	79	US-08-763-528A-2	Sequence 2, Appli
91	6	2.3	79	US-09-066-028-7	Sequence 7, Appli
92	6	2.3	79	US-09-066-028-8	Sequence 8, Appli
93	6	2.3	79	US-09-066-028-9	Sequence 9, Appli
94	6	2.3	79	US-09-066-028-10	Sequence 10, Appli
95	6	2.3	79	US-09-066-028-11	Sequence 11, Appli
96	6	2.3	79	US-09-335-325-7	Sequence 7, Appli
97	6	2.3	79	US-09-335-325-8	Sequence 8, Appli
98	6	2.3	79	US-09-335-325-9	Sequence 9, Appli
99	6	2.3	79	US-09-335-325-10	Sequence 10, Appli
100	6	2.3	79		

## ALIGNMENTS

```

RESULT 1
US-09-411-977-2
; Sequence 2, Application US/09411977
; Patent No. 6372473
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A.
; APPLICANT: Ruben, Steven M.
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Tissue Plasminogen Activator-Like Protease
; FILE REFERENCE: PR378P1
; CURRENT APPLICATION NUMBER: US/09/411,977
; CURRENT FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: 09/084,491
; EARLIER FILING DATE: 1998-05-27
; EARLIER APPLICATION NUMBER: 60/048,000
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-411-977-2

Query Match      100.0%; Score 263; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 5.2e-248;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MLANWQAFVLSNMLAEAYSGGCGFNDGNHLYREDOTSPAPGRCINMLDAOSGLASAP 60
      |||||||
DB      1 MLANWQAFVLSNMLAEAYSGGCGFNDGNHLYREDOTSPAPGRCINMLDAOSGLASAP 60

QY      61 VSGAGNSYCRNPEDPRGPMCYSGAGVPEKRPCCDLRCPETTSQALPAFTTEIOEAS 120
      |||||||
DB      61 VSGAGNSYCRNPEDPRGPMCYSGAGVPEKRPCCDLRCPETTSQALPAFTTEIOEAS 120

QY      121 EGPAGDEVQVAPANALPARSEAAAVQVIGISQVVMNSKEKKDLGTLGVIGITMMVI 180
      |||||||
DB      121 EGPAGDEVQVAPANALPARSEAAAVQVIGISQVVMNSKEKKDLGTLGVIGITMMVI 180

QY      181 IIAIGAGIILGYSTKRGDKLKEQDQKCEEMQRTILPLSAFNPICEIVDEKTVVYHT 240
      |||||||
DB      181 IIAIGAGIILGYSTKRGDKLKEQDQKCEEMQRTILPLSAFNPICEIVDEKTVVYHT 240

QY      241 SQTVPDPOEGSTPLMGQAGTGA 263
      |||||||
DB      241 SQTVPDPOEGSTPLMGQAGTGA 263

RESULT 2
US-08-766-982-1
; Sequence 1, Application US/08766982
; Patent No. 5948892
; GENERAL INFORMATION:
; APPLICANT: Wahl, Robert C.
; TITLE OF INVENTION: Analogs of Macrophage Stimulating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavenland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,982
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 716 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-766-982-1

Query Match      3.0%; Score 8; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      76 DPRGPWCY 83
      |||||||
DB      163 DPRGPWCY 170

```

```

RESULT 3
US-09-296-219-1
; Sequence 1, Application US/09296219
; Patent No. 6246560
; GENERAL INFORMATION:
; APPLICANT: Wahl, Robert C.
; TITLE OF INVENTION: Analogs of Macrophage Stimulating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavenland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/296,219
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 716 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-296-219-1

```

```

Query Match      3.0%; Score 8; DB 3; Length 716;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      76 DPRGPWCY 83
      |||||||
DB      163 DPRGPWCY 170

```

```
RESULT 4
US-09-198-452A-846
; Sequence 846, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 846
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; NAME/KEY: SITE
; LOCATION: 1...182
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-846

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 182;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDLGTLG 170
Db 42 KDLGTLG 48

RESULT 5
US-09-107-532A-5978
; Sequence 5978, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Denek
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 893-5007
; TELEFAX: (781) 893-8277
; INFORMATION FOR SEQ ID NO: 5978:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
MOLECULE TYPE: protein
;
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...204
; SEQUENCE DESCRIPTION: SEQ ID NO: 5978:
US-09-107-532A-5978

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 204;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SGLASAP 60
Db 98 SGLASAP 104

RESULT 6
US-09-198-452A-849
; Sequence 849, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 849
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-849

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 228;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KDLGTLG 170
Db 158 KDLGTLG 164

RESULT 7
US-09-252-991A-18308
; Sequence 18308, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18308
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18308

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 341;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 138 PARSEA 144  
|||  
Db 284 PARSEAA 290

## RESULT 8

US-09-107-532A-5053  
; Sequence 5053, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5053:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 353 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...353  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5053:  
US-09-107-532A-5053  
Query Match 2.7%; Score 7; DB 4; Length 353;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 IIAIGAG 187  
|||  
Db 207 IIAIGAG 213

## RESULT 9

US-09-107-532A-4004  
; Sequence 4004, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4004:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 356 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...356  
SEQUENCE DESCRIPTION: SEQ ID NO: 4004:  
US-09-107-532A-4004

Query Match 2.7%; Score 7; DB 4; Length 356;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 IIAIGAG 187  
|||  
Db 212 IIAIGAG 218

RESULT 10  
US-09-107-532A-6936  
; Sequence 6936, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII

US-09-107-532A-4004  
; Sequence 4004, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

```
/
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,532A
/ FILING DATE: 30-Jun-1998
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/085,598
/ FILING DATE: 14 May 1998
/ APPLICATION NUMBER: 60/051571
/ FILING DATE: July 2, 1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Denae
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-012
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/ INFORMATION FOR SEQ ID NO: 6936:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 356 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Enterococcus faecium
/
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (8) LOCATION 1...356
/ SEQUENCE DESCRIPTION: SEQ ID NO: 6936:
US-09-107-532A-6936

Query Match      2.7%; Score 7; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      187 GILGYS 193
      |||||
Db      144 GILGYS 150

RESULT 11
US-09-252-991A-29970
/ Sequence 29970, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 29970
/ LENGTH: 381
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29970

Query Match      2.7%; Score 7; DB 4; Length 381;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      71 RNPDEDP 77
      |||||
Db      127 RNPDEDP 133

RESULT 12
US-09-252-991A-28999
/ Sequence 28999, Application US/09252991A
```

```
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 28999
/ LENGTH: 413
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28999

Query Match      2.7%; Score 7; DB 4; Length 413;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      135 NALPARS 141
      |||||
Db      184 NALPARS 190

RESULT 13
US-08-009-075-4
/ Sequence 4, Application US/08009075
/ Patent No. 5300436
/ GENERAL INFORMATION:
/ APPLICANT: GOLDSTEIN, Menek
/ APPLICANT: WU, Jing
/ APPLICANT: FRIEDHOFF, David
/ APPLICANT: FRIEDHOFF, Arnold J.
/ TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE
/ TITLE OF INVENTION: HYDROXYLASE AND USES THEREOF
/ NUMBER OF SEQUENCES: 13
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: BROWDY AND NEIMARK
/ STREET: 419 Seventh Street, N.W., Suite 300
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20004
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/009,075
/ FILING DATE: 19930126
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: TOWNSEND, GUY K.
/ REGISTRATION NUMBER: 34,033
/ REFERENCE/DOCKET NUMBER: GOLDSTEIN-1A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-628-5197
/ TELEFAX: 202-737-3528
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 497 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-009-075-4

Query Match      2.7%; Score 7; DB 1; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 EAAAVOP 148  
|||||  
Db 415 EAAAVOP 421

## RESULT 14

US-08-467-822-27  
; Sequence 27, Application US/08467822

; Patent No. 5843460

; GENERAL INFORMATION:

; APPLICANT: Labigne, Agnes

; APPLICANT: Sauerbaum, Sebastien

; APPLICANT: Ferrero, Richard L.

; APPLICANT: Thiberge, Jean-Michel

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESS: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,822

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; INFORMATION FOR SEQ ID NO: 27:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/447,177

; FILING DATE: 19-MAY-1995

; CLASSIFICATION: 435

; APPLICATION NUMBER: US 08/432,697

; FILING DATE: 02-MAY-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 03495.0137-02000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 569 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-467-822-27

Query Match 2.7%; Score 7; DB 2; Length 569;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 EAGVPEK 93  
|||||  
Db 498 EAGVPEK 504

## RESULT 15

US-08-432-697-27

; Sequence 27, Application US/08432697  
; Patent No. 6248330

; GENERAL INFORMATION:

; APPLICANT: Labigne, Agnes

; APPLICANT: Sauerbaum, Sebastien

; APPLICANT: Ferrero, Richard L.

; APPLICANT: Thiberge, Jean-Michel

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESS: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/432,697

; FILING DATE: 02-MAY-1995

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 03495.0137-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 569 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-432-697-27

Query Match 2.7%; Score 7; DB 3; Length 569;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 EAGVPEK 93  
|||||  
Db 498 EAGVPEK 504

## RESULT 16

US-08-466-248-27

; Sequence 27, Application US/08466248

; Patent No. 6258359

; GENERAL INFORMATION:

; APPLICANT: Labigne, Agnes

; APPLICANT: Sauerbaum, Sebastien

; APPLICANT: Ferrero, Richard L.

; APPLICANT: Thiberge, Jean-Michel

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESS: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,248  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/447,177  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/432,697  
FILING DATE: 02-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0137-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 569 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-248-27

Query Match 2.7%; Score 7; DB 3; Length 569;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 EAGVPEK 93  
|||  
DB 498 EAGVPEK 504

RESULT 17  
US-08-403-866-7  
Sequence 7, Application US/08403866  
Patent No. 5643779  
GENERAL INFORMATION:  
APPLICANT: Ehrlich, Stanislaw  
APPLICANT: Godon, Jean-Jacques  
APPLICANT: Renault, Pierre  
TITLE OF INVENTION: Nucleic acid coding for an alpha-acetolactate  
TITLE OF INVENTION: synthase from Lactococcus and its applications  
NUMBER OF SEQUENCES: 16  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,866  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 20747/30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1600  
TELEFAX: (716) 263-1487  
TELEX: 978450 (WUT)

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Lactococcus lactis subsp. lactis  
INDIVIDUAL ISOLATE: ILV6  
US-08-403-866-7

Query Match 2.7%; Score 7; DB 1; Length 575;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 LLAAYG 21  
|||  
DB 504 LLAAYG 510

RESULT 18  
US-08-532-547-7  
Sequence 7, Application US/08532547  
Patent No. 5861543  
GENERAL INFORMATION:  
APPLICANT: LAMBERT, BART  
APPLICANT: JANSENS, STEFAN  
APPLICANT: VAN AUDENHOVE, KATRIEN  
APPLICANT: PEERDEN, MARINX  
APPLICANT: VAN RIE, JEROEN  
APPLICANT: VAN AARSSEN, ROEL  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,547  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-109P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-532-547-7

Query Match 2.7%; Score 7; DB 2; Length 625;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 ASAPVSG 63  
|||

Db 482 ASAPVSG 488

```
RESULT 19
US-08-532-547-9
; Sequence 9, Application US/08532547
; Patent No. 5861543
; GENERAL INFORMATION:
; APPLICANT: LAMBERT, BART
; APPLICANT: JANSSENS, STEFAN
; APPLICANT: VAN AUDENHOVE, KATRIEN
; APPLICANT: PEERROEN, MARINX
; APPLICANT: VAN RIE, JEROEN
; APPLICANT: VAN AARSEN, ROEL
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,547
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-109P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 625 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-532-547-9

Query Match 2.7%; Score 7; DB 2; Length 625;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 57 ASAPVSG 63  
|||||  
Db 482 ASAPVSG 488

```
RESULT 20
US-09-019-809-7
; Sequence 7, Application US/09019809
; Patent No. 6143550
; GENERAL INFORMATION:
; APPLICANT: LAMBERT, BART
; APPLICANT: JANSSENS, STEFAN
; APPLICANT: VAN AUDENHOVE, KATRIEN
; APPLICANT: PEERROEN, MARINX
; APPLICANT: VAN RIE, JEROEN
; APPLICANT: VAN AARSEN, ROEL
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,809
; FILING DATE: 02-FEB-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-135P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 625 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-019-809-7

Query Match 2.7%; Score 7; DB 3; Length 625;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 57 ASAPVSG 63  
|||||  
Db 482 ASAPVSG 488

```
RESULT 21
US-09-019-809-9
; Sequence 9, Application US/09019809
; Patent No. 6143550
; GENERAL INFORMATION:
; APPLICANT: LAMBERT, BART
; APPLICANT: JANSSENS, STEFAN
; APPLICANT: VAN AUDENHOVE, KATRIEN
; APPLICANT: PEERROEN, MARINX
; APPLICANT: VAN RIE, JEROEN
; APPLICANT: VAN AARSEN, ROEL
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR
; TITLE OF INVENTION: INSECTICIDAL PROTEINS.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,809
; FILING DATE: 02-FEB-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
```



REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-019-809-9

Query Match 2.7%; Score 7; DB 3; length 625;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ASAPVSG 63  
Db 482 ASAPVSG 488

RESULT 22  
US-09-471-177-7  
Sequence 7, Application US/09471177  
Patent No. 6448226  
GENERAL INFORMATION:  
APPLICANT: LAMBERT, BART  
APPLICANT: JANSSENS, STEFAN  
APPLICANT: VAN AUDENHOVE, KATRINEN  
APPLICANT: PEPEROEN, MARINX  
APPLICANT: VAN RIE, JEROEN  
APPLICANT: VAN AARSEN, ROEL  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/471.177  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/019.809  
FILING DATE: 02-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-471-177-7

Query Match 2.7%; Score 7; DB 4; length 625;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 57 ASAPVSG 63  
Db 482 ASAPVSG 488

RESULT 23  
US-09-471-177-9  
Sequence 9, Application US/09471177  
Patent No. 6448226  
GENERAL INFORMATION:  
APPLICANT: LAMBERT, BART  
APPLICANT: JANSSENS, STEFAN  
APPLICANT: VAN AUDENHOVE, KATRINEN  
APPLICANT: PEPEROEN, MARINX  
APPLICANT: VAN RIE, JEROEN  
APPLICANT: VAN AARSEN, ROEL  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/471.177  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/019.809  
FILING DATE: 02-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 625 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-471-177-9

Query Match 2.7%; Score 7; DB 4; length 625;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ASAPVSG 63  
Db 482 ASAPVSG 488

RESULT 24  
US-09-946-678-2  
Sequence 2, Application US/09946678  
Patent No. 6541236  
GENERAL INFORMATION:  
APPLICANT: ITO, Kotaro  
APPLICANT: UMITSUKI, Genryou  
APPLICANT: KOTAMA, Yasuji

;; TITLE OF INVENTION: Protein Having Glutaminase Activity and Gene Encoding the Same  
;; FILE REFERENCE: 0283-0158P  
;; CURRENT APPLICATION NUMBER: US/09/946,678  
;; CURRENT FILING DATE: 2001-09-06  
;; PRIOR APPLICATION NUMBER: JP 2000-270371  
;; PRIOR FILING DATE: 2000-09-06  
;; NUMBER OF SEQ ID NOS: 27  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 2  
;; LENGTH: 684  
;; TYPE: PRT  
;; ORGANISM: Cryptococcus nodaeensis  
US-09-946-678-2

Query Match 2.7%; Score 7; DB 4; Length 684;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 DAOSGLA 57  
Db 411 DAOSGLA 417

RESULT 25  
US-08-469-537A-72  
; Sequence 72, Application US/08469537A  
; Patent No. 5843749  
; GENERAL INFORMATION:  
; APPLICANT: Maisompierre, et al.  
; TITLE OF INVENTION: EHK AND ROR TYROSINE  
; TITLE OF INVENTION: KINASES  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
; STREET: 777 Old Saw Mill River Road  
; CITY: Tarrytown  
; STATE: NY  
; COUNTRY: U.S.A.  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,537A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 08/406,247  
; FILING DATE: 17-MAR-1995  
; APPLICATION NUMBER: USSN 08/144,992  
; FILING DATE: 28-OCT-1993  
; APPLICATION NUMBER: USSN 07/736,559  
; FILING DATE: 26-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kempster, Ph.D., Gail M  
; REGISTRATION NUMBER: 32,143  
; REFERENCE/DOCKET NUMBER: REG 070C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 914-345-7400  
; TELEFAX: 914-345-7721  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 72:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 800 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-469-537A-72

Query Match 2.7%; Score 7; DB 2; Length 800;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 HSYCRNP 73  
Db 222 HSYCRNP 228

RESULT 26  
US-08-469-537A-78  
; Sequence 78, Application US/08469537A  
; Patent No. 5843749  
; GENERAL INFORMATION:  
; APPLICANT: Maisompierre, et al.  
; TITLE OF INVENTION: EHK AND ROR TYROSINE  
; TITLE OF INVENTION: KINASES  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
; STREET: 777 Old Saw Mill River Road  
; CITY: Tarrytown  
; STATE: NY  
; COUNTRY: U.S.A.  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,537A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 08/406,247  
; FILING DATE: 17-MAR-1995  
; APPLICATION NUMBER: USSN 08/144,992  
; FILING DATE: 28-OCT-1993  
; APPLICATION NUMBER: USSN 07/736,559  
; FILING DATE: 26-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kempster, Ph.D., Gail M  
; REGISTRATION NUMBER: 32,143  
; REFERENCE/DOCKET NUMBER: REG 070C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 914-345-7400  
; TELEFAX: 914-345-7721  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 78:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 800 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-469-537A-78

Query Match 2.7%; Score 7; DB 2; Length 800;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 HSYCRNP 73  
Db 222 HSYCRNP 228

RESULT 27  
US-09-269-861A-8  
; Sequence 8, Application US/09269861A  
; Patent No. 6468775  
; GENERAL INFORMATION:  
; APPLICANT: Ankenbauer, Maltraud  
; APPLICANT: Markau, Ursula  
; APPLICANT: Svetlichny, Vitaly

APPLICANT: Schmitz-Aghegarian, Gudrun  
APPLICANT: Reiser, Astrid  
APPLICANT: Angerer, Bernhard  
APPLICANT: Ebenbichler, Christine  
APPLICANT: Laue, Frank  
APPLICANT: Bonch-Osmolovskaya, Elizaveta  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE FROM CARBOXYDOTHERMUS HYDROGENOFORMAN  
FILE REFERENCE: 4494  
CURRENT APPLICATION NUMBER: US/09/269,861A  
CURRENT FILING DATE: 1999-11-22  
PRIOR APPLICATION NUMBER: PCT/EP97/05391  
PRIOR FILING DATE: 1997-10-01  
PRIOR APPLICATION NUMBER: EP/96115873.0  
PRIOR FILING DATE: 1996-10-03  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8  
LENGTH: 831  
TYPE: PRT  
ORGANISM: Carboxydothermus hydrogenoformans  
US-09-269-861A-8

Query Match 2.7%; Score 7; DB 4; Length 831;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 231 VDEKTV 237  
Db 140 VDEKTV 146

RESULT 28  
US-08-469-537A-105  
Sequence 105, Application US/08469537A  
Patent No. 5843749  
GENERAL INFORMATION:  
APPLICANT: Maisompierre, et al.  
TITLE OF INVENTION: EHK AND FOR TYROSINE  
TITLE OF INVENTION: KINASES  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PasteSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,537A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/406,247  
FILING DATE: 17-MAR-1995  
APPLICATION NUMBER: USSN 08/144,992  
FILING DATE: 28-OCT-1993  
APPLICATION NUMBER: USSN 07/736,559  
FILING DATE: 26-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempier, Ph.D., Ga11 M  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 0700C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
TELEX:  
INFORMATION FOR SEQ ID NO: 105:  
SEQUENCE CHARACTERISTICS:

LENGTH: 937 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Human ROR1  
LOCATION: 1...937  
OTHER INFORMATION:  
US-08-469-537A-105

Query Match 2.7%; Score 7; DB 2; Length 937;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 67 HSYCRNP 73  
Db 359 HSYCRNP 365

RESULT 29  
US-09-002-285-72  
Sequence 72, Application US/09002285  
Patent No. 6369213  
GENERAL INFORMATION:  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Wicker, Carol  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Walz, Michelle  
APPLICANT: Stockhoff, Brian  
APPLICANT: Muller-Cohn, Judy  
TITLE OF INVENTION: Toxins Active Against Pests  
NUMBER OF SEQUENCES: 105  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Salivanchik, Lloyd & Salivanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/002,285  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/886,615  
FILING DATE: 1-JUL-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/674,002  
FILING DATE: 1-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M.  
REGISTRATION NUMBER: 39,355  
REFERENCE/DOCKET NUMBER: MA-701C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (352) 375-8100  
TELEFAX: (352) 372-5800  
INFORMATION FOR SEQ ID NO: 72:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1156 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-002-285-72  
Query Match 2.7%; Score 7; DB 4; Length 1156;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 ASAPVSG 63  
|||  
Db 523 ASAPVSG 529

RESULT 30  
US-09-589-477-72  
; Sequence 72, Application US/09589477  
; Patent No. 6570005  
; GENERAL INFORMATION:  
; APPLICANT: Schmedf, H. Ernest  
; APPLICANT: Wicker, Carol  
; APPLICANT: Narva, Kenneth E.  
; APPLICANT: Walz, Michelle  
; APPLICANT: Stockhoff, Brian  
; TITLE OF INVENTION: Toxins Active Against Pests  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/589,477  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/886,615  
; FILING DATE: 1-JUL-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/674,002  
; FILING DATE: 1-JUL-1996  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sanders, Jay M.  
; REGISTRATION NUMBER: 39,355  
; REFERENCE/DOCKET NUMBER: MA-701C1C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (352) 375-8100  
; TELEFAX: (352) 372-5800  
; INFORMATION FOR SEQ ID NO: 72:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1156 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-589-477-72

Query Match 2.7%; Score 7; DB 4; Length 1156;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 ASAPVSG 63  
|||  
Db 523 ASAPVSG 529

; GENERAL INFORMATION:  
; APPLICANT: LAMBERT, BART  
; APPLICANT: JANSSENS, STEFAN  
; APPLICANT: VAN AUDENHOVE, KATRIEN  
; APPLICANT: PEFEROEN, MARINX  
; APPLICANT: VAN RIE, JEROEN  
; APPLICANT: VAN AARSSEN, ROEL  
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
; INSECTICIDAL PROTEINS.  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/532,547  
; FILING DATE: 06-DEC-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SVENSSON, LEONARD R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 2121-109P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1157 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-532-547-5

Query Match 2.7%; Score 7; DB 2; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 ASAPVSG 63  
|||  
Db 523 ASAPVSG 529

RESULT 32  
US-08-379-656B-5  
; Sequence 5, Application US/08379656B  
; Patent No. 5885571  
; GENERAL INFORMATION:  
; APPLICANT: Lambert Bart  
; APPLICANT: Janssens, Stefan  
; APPLICANT: Van Audenove, Katrien  
; APPLICANT: Peferoen, Marinix  
; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND  
; THEIR INSECTICIDAL PROTEINS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 810 Gatehouse Road, Suite 500 East  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/379,656B  
FILING DATE: 23-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP93/01820  
FILING DATE: 12-JULY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 93400949.9  
FILING DATE: 09-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92402358.8  
FILING DATE: 27-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-104P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1157 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-379-656B-5

Query Match 2.7%; Score 7; DB 2; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 ASAPVSG 63  
Db 523 ASAPVSG 529

RESULT 33  
US-08-455-838-5  
Sequence 5, Application US/08455838  
Patent No. 6028246  
GENERAL INFORMATION:  
APPLICANT: Lambert Bart  
APPLICANT: Jansens, Stefan  
APPLICANT: Van Audenhove, Katrien  
APPLICANT: Peferoen, Marnix  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND  
TITLE OF INVENTION: THEIR INSECTICIDAL PROTEINS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,838  
FILING DATE: 31-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP93/01820  
FILING DATE: 12-JULY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 93400949.9  
FILING DATE: 09-APR-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 92402358.8  
FILING DATE: 27-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1157 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-455-838-5

Query Match 2.7%; Score 7; DB 3; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 ASAPVSG 63  
Db 523 ASAPVSG 529

RESULT 34  
US-09-019-809-5  
Sequence 5, Application US/09019809  
Patent No. 6143550  
GENERAL INFORMATION:  
APPLICANT: LAMBERT, BART  
APPLICANT: JANSSENS, STEFAN  
APPLICANT: VAN AUDENHOVE, KATRIEN  
APPLICANT: PEFEROEN, MARNIX  
APPLICANT: VAN RIE, JEROEN  
APPLICANT: VAN AARSEN, ROEL  
TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR  
TITLE OF INVENTION: INSECTICIDAL PROTEINS.  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/019,809  
FILING DATE: 02-FEB-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-135P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1157 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-019-809-5

Query Match 2.7%; Score 7; DB 3; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ASAPVSG 63  
Db 523 ASAPVSG 529

RESULT 35

US-09-471-177-5

; Sequence 5, Application US/09471177

; Patent No. 6448226

; GENERAL INFORMATION:

; APPLICANT: LAMBERT, BART

; APPLICANT: JANSSEN, STEFAN

; APPLICANT: VAN AUDENHOVE, KATHIEN

; APPLICANT: PEPEROEN, MARINIX

; APPLICANT: VAN RIE, JEROEN

; TITLE OF INVENTION: NEW BACILLUS THURINGIENSIS STRAINS AND THEIR

; TITLE OF INVENTION: INSECTICIDAL PROTEINS.

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP

; STREET: P.O. Box 747

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/471,177

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/019,809

; FILING DATE: 02-FEB-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: SVENSSON, LEONARD R.

; REGISTRATION NUMBER: 30,330

; REFERENCE/DOCKET NUMBER: 2121-135P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 205-8000

; TELEFAX: (703) 205-8050

; TELEX: 248345

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1157 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-471-177-5

Query Match 2.7%; Score 7; DB 4; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ASAPVSG 63  
Db 523 ASAPVSG 529

RESULT 36

US-09-252-991A-32960

; Sequence 32960, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 32960

; LENGTH: 1247

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; US-09-252-991A-32960

Query Match 2.7%; Score 7; DB 4; Length 1247;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 RSEAAV 146  
Db 920 RSEAAV 926

RESULT 37

US-09-192-012-8

; Sequence 8, Application US/09192012A

; Patent No. 6475784

; GENERAL INFORMATION:

; APPLICANT: Papkoef, Jackie

; APPLICANT: Megabios Corporation

; APPLICANT: Pfizer, Inc.

; TITLE OF INVENTION: Inhibition of Angiogenesis by Delivery of Nucleic Acids

; TITLE OF INVENTION: Encoding Anti-Angiogenesis Polypeptides

; FILE REFERENCE: 018484-00010US

; CURRENT APPLICATION NUMBER: US/09/192,012A

; CURRENT FILING DATE: 1998-11-13

; EARLIER APPLICATION NUMBER: US 60/066,020

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 8

; LENGTH: 7

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:consensus

; US-09-192-012-8

Query Match 2.3%; Score 6; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
Db 2 YCRNPD 7

RESULT 38

US-08-267-092A-12

; Sequence 12, Application US/08267092A

; Patent No. 5599686

; GENERAL INFORMATION:

; APPLICANT: Defeo-Jones, Deborah

; APPLICANT: Garsky, Victor M.

; APPLICANT: Jones, Raymond E.

; APPLICANT: Oliff, Allen I.

; TITLE OF INVENTION: NOVEL PEPTIDES

; NUMBER OF SEQUENCES: 65

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David A. Multhard

```
STREET: 126 E. Lincoln Avenue, P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/267,092A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mulhard, David A.
REGISTRATION NUMBER: 35,297
REFERENCE/DOCKET NUMBER: 19253
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3903
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-267-092A-12

Query Match      2.3%; Score 6; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 258 AGTPGA 263  
|||||  
Db 3 AGTPGA 8

```
RESULT 39
US-08-468-161-12
Sequence 12, Application US/08468161
GENERAL INFORMATION:
APPLICANT: Defeo-Jones, Deborah
APPLICANT: Feng, Dong-Mei
APPLICANT: Garsky, Victor M.
APPLICANT: Jones, Raymond E.
APPLICANT: Oliff, Allen I.
TITLE OF INVENTION: NOVEL PEPTIDES
NUMBER OF SEQUENCES: 146
CORRESPONDENCE ADDRESS:
ADDRESSER: DAVID A. MUTHARD
STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
CITY: RAHWAY
STATE: NEW JERSEY
COUNTRY: U.S.A.
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,161
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mulhard, David A.
REGISTRATION NUMBER: 35,297
```

```
REFERENCE/DOCKET NUMBER: 192531B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908)594-3903
TELEFAX: (908)594-4720
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-468-161-12

Query Match      2.3%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 258 AGTPGA 263  
|||||  
Db 3 AGTPGA 8

```
RESULT 40
PCT-US95-08156-12
Sequence 12, Application PC/TUS9508156
GENERAL INFORMATION:
APPLICANT: Defeo-Jones, Deborah
APPLICANT: Feng, Dong-Mei
APPLICANT: Garsky, Victor M.
APPLICANT: Jones, Raymond E.
APPLICANT: Oliff, Allen I.
TITLE OF INVENTION: NOVEL PEPTIDES
NUMBER OF SEQUENCES: 146
CORRESPONDENCE ADDRESS:
ADDRESSER: DAVID A. MUTHARD
STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
CITY: RAHWAY
STATE: NEW JERSEY
COUNTRY: U.S.A.
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08156
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mulhard, David A.
REGISTRATION NUMBER: 35,297
REFERENCE/DOCKET NUMBER: 192531B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908)594-4720
TELEFAX: (908)594-3903
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
PCT-US95-08156-12

Query Match      2.3%; Score 6; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
```

```
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      258 ACTPGA 263
      |||||
Db      3 ACTPGA 8

RESULT 41
US-09-348-953-4
; Sequence 4, Application US/09348953
; Patent No. 6538103
; GENERAL INFORMATION:
; APPLICANT: Ji, Richard Wei-Dong
; APPLICANT: Trail, Pamela A.
; TITLE OF INVENTION: LYSINE BINDING FRAGMENTS OF ANGIOSTATIN
; FILE REFERENCE: DB1Sequences
; CURRENT APPLICATION NUMBER: US/09/348,953
; PRIOR FILING DATE: 1997-07-07
; PRIOR FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 38
; TYPE: PRT
; ORGANISM: human
US-09-348-953-4

Query Match      2.3%; Score 6; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      69 YCRNPD 74
      |||||
Db      16 YCRNPD 21

RESULT 42
US-09-227-357-589
; Sequence 589, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; PRIOR FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
```

```
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 589
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-589

Query Match      2.3%; Score 6; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      80 PWCYVS 85
      |||||
Db      32 PWCYVS 37

RESULT 43
US-09-348-953-6
; Sequence 6, Application US/09348953
; Patent No. 6538103
; GENERAL INFORMATION:
; APPLICANT: Ji, Richard Wei-Dong
; APPLICANT: Trail, Pamela A.
; TITLE OF INVENTION: LYSINE BINDING FRAGMENTS OF ANGIOSTATIN
; FILE REFERENCE: DB1Sequences
; CURRENT APPLICATION NUMBER: US/09/348,953
; PRIOR FILING DATE: 1999-07-07
; PRIOR FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
```



```

; LENGTH: 46
; TYPE: PRT
; ORGANISM: human
US-09-348-953-6

Query Match
Best Local Similarity 100.0%; Score 6; DB 4; Length 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 69 YCRNPD 74
Db 25 YCRNPD 30

RESULT 44
US-09-227-357-588
; Sequence 588, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
```

```

; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 588
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-588

Query Match
Best Local Similarity 100.0%; Score 6; DB 4; Length 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 PRCYVS 85
Db 46 PRCYVS 51

RESULT 45
US-08-630-915A-215
; Sequence 215, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 215:
```

SEQUENCE CHARACTERISTICS:  
LENGTH: 59 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-215

Query Match 2.3%; Score 6; DB 4; Length 59;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 VDEKTV 236  
|||||  
DB 29 VDEKTV 34

RESULT 46  
US-09-107-532A-6727  
; Sequence 6727, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 6727:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 66 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...66  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6727:  
US-09-107-532A-6727

Query Match 2.3%; Score 6; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 183 AKGAGI 188

DB 45 AIGAGI 50  
|||||

RESULT 47  
US-09-107-532A-4213  
; Sequence 4213, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 4213:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 73 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...73  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4213:  
US-09-107-532A-4213

Query Match 2.3%; Score 6; DB 4; Length 73;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 EAGVPE 92  
|||||  
DB 40 EAGVPE 45

RESULT 48  
US-08-612-788-13  
; Sequence 13, Application US/08612788  
; Patent No. 5837682  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah  
; APPLICANT: O'Reilly, Michael  
; APPLICANT: Cao, Yihai  
; APPLICANT: Sim, B. Kim Lee

TITLE OF INVENTION: Angiotensin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
IMMEDIATE SOURCE:  
CLONE: K2  
US-08-612-788-13

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
| | | | |  
DB 49 YCRNPD 54

RESULT 49  
US-08-612-788-14  
Sequence 14, Application US/08612788  
Patent No. 5837682  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiotensin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Rhesus monkey  
IMMEDIATE SOURCE:  
CLONE: K2  
US-08-612-788-14

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
| | | | |  
DB 49 YCRNPD 54

RESULT 50  
US-08-612-788-15  
Sequence 15, Application US/08612788  
Patent No. 5837682  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiotensin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids

TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Porcine  
IMMEDIATE SOURCE:  
CLONE: K2  
US-08-612-788-15

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred.No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YCRNPD 74  
Db 49 YCRNPD 54

RESULT 51

US-08-612-788-16  
; Sequence 16, Application US/08612788  
; Patent No. 5837682  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah  
; APPLICANT: O'Reilly, Micheal  
; APPLICANT: Cao, Yihai  
; APPLICANT: Sam, B. Kim Lee  
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.  
; ZIP: 30303-1769  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/612,788  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren, William L.  
; REGISTRATION NUMBER: 36,714  
; REFERENCE/DOCKET NUMBER: 05213-0126  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-818-3700  
; TELEFAX: 404-818-3799  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Bovine  
; IMMEDIATE SOURCE:  
; CLONE: K2  
US-08-612-788-16

Query Match 2.3%; Score 6; DB 2; Length 78;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YCRNPD 74  
Db 49 YCRNPD 54

RESULT 52

US-08-612-788-17  
; Sequence 17, Application US/08612788  
; Patent No. 5837682  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah  
; APPLICANT: O'Reilly, Micheal  
; APPLICANT: Cao, Yihai  
; APPLICANT: Sam, B. Kim Lee  
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.  
; ZIP: 30303-1769  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/612,788  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren, William L.  
; REGISTRATION NUMBER: 36,714  
; REFERENCE/DOCKET NUMBER: 05213-0126  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-818-3700  
; TELEFAX: 404-818-3799  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Murine  
; IMMEDIATE SOURCE:  
; CLONE: K3  
US-08-612-788-17

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred.No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YCRNPD 74  
Db 49 YCRNPD 54

RESULT 53

US-08-612-788-18  
; Sequence 18, Application US/08612788  
; Patent No. 5837682  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, M. Judah

APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: K3  
US-08-612-788-18

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred.No. 2,1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
Db 49 YCRNPD 54

RESULT 54  
US-08-612-788-19  
Sequence 19, Application US/08612788  
Patent No. 5837682  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Rhesus monkey  
IMMEDIATE SOURCE:  
CLONE: K3  
US-08-612-788-19

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred.No. 2,1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
Db 49 YCRNPD 54

RESULT 55  
US-08-612-788-20  
Sequence 20, Application US/08612788  
Patent No. 5837682  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Porcine  
IMMEDIATE SOURCE:  
CLONE: K3  
US-08-612-788-20

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YCRNPD 74  
Db 49 YCRNPD 54

RESULT 56  
US-08-612-788-21  
Sequence 21, Application US/08612788  
Patent No. 5837682  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Michael  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612.788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Bovine  
IMMEDIATE SOURCE:  
CLONE: K3

US-08-612-788-21

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YCRNPD 74  
Db 49 YCRNPD 54

RESULT 57  
US-08-612-788-22  
Sequence 22, Application US/08612788  
Patent No. 5837682  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Michael  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612.788  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Murine  
IMMEDIATE SOURCE:  
CLONE: K4  
US-08-612-788-22

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YCRNPD 74  
Db 49 YCRNPD 54

RESULT 58  
US-08-612-788-23  
Sequence 23, Application US/08612788

```
/ Patent No. 5837682
/ GENERAL INFORMATION:
/ APPLICANT: Folkman, M. Judah
/ APPLICANT: O'Reilly, Michael
/ APPLICANT: Cao, Yihai
/ APPLICANT: Sim, B. Kim Lee
/ TITLE OF INVENTION: Angiostatin Fragments and Method of Use
/ NUMBER OF SEQUENCES: 45
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Jones & Askew
/ STREET: 191 Peachtree Street, 37th Floor
/ CITY: Atlanta
/ STATE: Georgia
/ COUNTRY: U.S.
/ ZIP: 30303-1769
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/612,788
/ FILING DATE:
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Warren, William L.
/ REGISTRATION NUMBER: 36,714
/ REFERENCE/DOCKET NUMBER: 05213-0126
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 404-818-3700
/ TELEFAX: 404-818-3799
/ INFORMATION FOR SEQ ID NO: 23:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 78 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: N-terminal
/ ORIGINAL SOURCE:
/ ORGANISM: Homo sapiens
/ IMMEDIATE SOURCE:
/ CLONE: K4
/ US-08-612-788-23

Query Match          2.3%; Score 6; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      69 YCRNPD 74
      |||||
Db      49 YCRNPD 54

RESULT 59
US-08-763-528A-3
/ Sequence 3, Application US/08763528A
/ Patent No. 5854221
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yihai
/ APPLICANT: Folkman, M. Judah
/ TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor
/ TITLE OF INVENTION: and Method of Use
/ NUMBER OF SEQUENCES: 6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Jones & Askew, LLP
/ STREET: 191 Peachtree Street, 37th Floor
/ CITY: Atlanta
/ STATE: Georgia
/ COUNTRY: US
/ ZIP: 30303
```

```
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/763,528A
/ FILING DATE: 12-DEC-1996
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Warren, William L.
/ REGISTRATION NUMBER: 36,714
/ REFERENCE/DOCKET NUMBER: 05940-0251
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 404-818-3700
/ TELEFAX: 404-818-3799
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 78 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/ FEATURE:
/ NAME/KEY: Protein
/ LOCATION: 1..78
/ OTHER INFORMATION: /note= "K-ingle 2 - Figure 3"
/ US-08-763-528A-3

Query Match          2.3%; Score 6; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      69 YCRNPD 74
      |||||
Db      49 YCRNPD 54

RESULT 60
US-08-763-528A-4
/ Sequence 4, Application US/08763528A
/ Patent No. 5854221
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yihai
/ APPLICANT: Folkman, M. Judah
/ TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor
/ TITLE OF INVENTION: and Method of Use
/ NUMBER OF SEQUENCES: 6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Jones & Askew, LLP
/ STREET: 191 Peachtree Street, 37th Floor
/ CITY: Atlanta
/ STATE: Georgia
/ COUNTRY: US
/ ZIP: 30303
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/763,528A
/ FILING DATE: 12-DEC-1996
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Warren, William L.
/ REGISTRATION NUMBER: 36,714
/ REFERENCE/DOCKET NUMBER: 05940-0251
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 404-818-3700
```

TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..78  
OTHER INFORMATION: /note= "Xringle 3 - Figure 3"  
US-08-763-528A-4

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
|||||  
DB 49 YCRNPD 54

RESULT 61  
US-08-763-528A-5  
Sequence 5, Application US/08763528A  
Patent No. 5854221  
GENERAL INFORMATION:  
APPLICANT: Cao, Yihai  
APPLICANT: Folkman, M. Judah  
TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor  
TITLE OF INVENTION: and Method of Use  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/763,528A  
FILING DATE: 12-DEC-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05940-0251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..78  
OTHER INFORMATION: /note= "Xringle 4 - Figure 3"

US-08-763-528A-5

Query Match 2.3%; Score 6; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
|||||  
DB 49 YCRNPD 54

RESULT 62  
US-09-066-028-13  
Sequence 13, Application US/09066028  
Patent No. 6024688  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
IMMEDIATE SOURCE:  
CLONE: K2  
US-09-066-028-13

Query Match 2.3%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
|||||  
DB 49 YCRNPD 54



```
RESULT 63
US-09-066-028-14
; Sequence 14, Application US/09066028
; Patent No. 6024688
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/612,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0126
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Rhesus monkey
; IMMEDIATE SOURCE:
; CLONE: K2
; US-09-066-028-14

Query Match      2.3%; Score 6; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 64
US-09-066-028-15
; Sequence 15, Application US/09066028
; Patent No. 6024688
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45

RESULT 65
US-09-066-028-16
; Sequence 16, Application US/09066028
; Patent No. 6024688
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

CORRESPONDENCE ADDRESS:
ADDRESSEE: Jones & Askew
STREET: 191 Peachtree Street, 37th Floor
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30303-1769
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/612,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren, William L.
REGISTRATION NUMBER: 36,714
REFERENCE/DOCKET NUMBER: 05213-0126
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Porcine
IMMEDIATE SOURCE:
CLONE: K2
; US-09-066-028-15

Query Match      2.3%; Score 6; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Bovine  
IMMEDIATE SOURCE:  
CLONE: K2  
US-09-066-028-16

Query Match 2.3%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred.No.2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
|||||  
DB 49 YCRNPD 54

RESULT 66  
US-09-066-028-17  
Sequence 17, Application US/09066028  
GENERAL INFORMATION:  
PATENT NO. 6024688  
APPLICANT: Folkman, M. Judah  
APPLICANT: Cao, Yihai  
APPLICANT: O'Reilly, Micheal  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714

REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Murine  
IMMEDIATE SOURCE:  
CLONE: K3  
US-09-066-028-17

Query Match 2.3%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred.No.2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
|||||  
DB 49 YCRNPD 54

RESULT 67  
US-09-066-028-18  
Sequence 18, Application US/09066028  
GENERAL INFORMATION:  
PATENT NO. 6024688  
APPLICANT: Folkman, M. Judah  
APPLICANT: Cao, Yihai  
APPLICANT: O'Reilly, Micheal  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: K3  
US-09-066-028-18

Query Match 2.3%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
Db 49 YCRNPD 54

## RESULT 68

US-09-066-028-19  
Sequence 19, Application US/09066028  
Patent No. 6024688  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Rhesus monkey  
IMMEDIATE SOURCE:  
CLONE: K3  
US-09-066-028-19

Query Match 2.3%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 YCRNPD 74  
Db 49 YCRNPD 54

## RESULT 69

US-09-066-028-20  
Sequence 20, Application US/09066028  
Patent No. 6024688  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
APPLICANT: O'Reilly, Micheal  
APPLICANT: Cao, Yihai  
APPLICANT: Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Porcine  
IMMEDIATE SOURCE:  
CLONE: K3  
US-09-066-028-20

Query Match 2.3%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
Db 49 YCRNPD 54

## RESULT 70

US-09-066-028-21  
Sequence 21, Application US/09066028  
Patent No. 6024688

```

; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/612,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0126
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; IMMEDIATE SOURCE:
; CLONE: K3
;
US-09-066-028-21
;
Query Match      2.3%; Score 6; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY      69 YCRNPD 74      |||||
Db      49 YCRNPD 54
;
RESULT 71
US-09-066-028-22
; Sequence 22, Application US/09066028
; Patent No. 6024688
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
;

```

```

; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/612,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0126
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Murine
; IMMEDIATE SOURCE:
; CLONE: K4
;
US-09-066-028-22
;
Query Match      2.3%; Score 6; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY      69 YCRNPD 74      |||||
Db      49 YCRNPD 54
;
RESULT 72
US-09-066-028-23
; Sequence 23, Application US/09066028
; Patent No. 6024688
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,028
; FILING DATE:
;

```

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/612,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: K4  
US-09-066-028-23

Query Match 2.3%; Score 6; DB 3; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YCRNDP 74  
|||||  
Db 49 YCRNDP 54

RESULT 73  
US-09-335-325-13  
Sequence 13, Application US/09335325  
Patent No. 6521439  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
O'Reilly, Micheal  
Cao, Yihai  
Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/335,325  
FILING DATE: 17-Jun-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
IMMEDIATE SOURCE:  
CLONE: K2  
US-09-335-325-13

Query Match 2.3%; Score 6; DB 4; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YCRNDP 74  
|||||  
Db 49 YCRNDP 54

RESULT 74  
US-09-335-325-14  
Sequence 14, Application US/09335325  
Patent No. 6521439  
GENERAL INFORMATION:  
APPLICANT: Folkman, M. Judah  
O'Reilly, Micheal  
Cao, Yihai  
Sim, B. Kim Lee  
TITLE OF INVENTION: Angiostatin Fragments and Method of Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/335,325  
FILING DATE: 17-Jun-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,788  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0126  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:  
ORGANISM: Rhesus monkey  
IMMEDIATE SOURCE:  
CLONE: K2  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-335-325-14

Query Match 2.3%; Score 6; DB 4; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YCRNPD 74  
|||||  
Db 49 YCRNPD 54

RESULT 75  
US-09-335-325-15  
Sequence 15, Application US/09335325  
Patent No. 6521439

GENERAL INFORMATION:

APPLICANT: Folkman, M. Judah  
O'Reilly, Micheal  
Gao, Yihai  
Sim, B. Kim Lee

TITLE OF INVENTION: Angiostatin Fragments and Method of Use

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSER: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/335.325

FILING DATE: 17-Jun-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/612,788

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Warren, William D.

REGISTRATION NUMBER: 36,714

REFERENCE/DOCKET NUMBER: 05213-0126

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 78 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:

ORGANISM: Porcine

IMMEDIATE SOURCE:

CLONE: K2

SEQUENCE DESCRIPTION: SEQ ID NO: 15:  
US-09-335-325-15

QY 69 YCRNPD 74  
|||||  
Db 49 YCRNPD 54

Search completed: November 25, 2003, 13:42:13  
Job time : 22 secs

Query Match 2.3%; Score 6; DB 4; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;